Winter 1-15-2021

Testing the Common-Mechanisms Theory of False Hearing and False Memory: The Roles of Executive Functioning and Inhibitory Control

Eric Failes
Washington University in St. Louis

Recommended Citation

Failes, Eric, "Testing the Common-Mechanisms Theory of False Hearing and False Memory: The Roles of Executive Functioning and Inhibitory Control" (2021). *Arts & Sciences Electronic Theses and Dissertations*. 2314.
https://openscholarship.wustl.edu/art_sci_etds/2314

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychological & Brain Sciences

Dissertation Examination Committee:
Mitchell S. Sommers, Chair
Mark A. McDaniel
Jonathan E. Peelle
Henry L. Roediger, III
Kristin J. Van Engen

Testing the Common-Mechanisms Theory of False Hearing and False Memory: The Roles of Executive Functioning and Inhibitory Control

by
Eric Failes

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

January 2021
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to begin by thanking each of my dissertation committee members: Mitch Sommers, Kristin Van Engen, Roddy Roediger, Jonathan Peelle, and Mark McDaniel. As usual, I decided to make things hard on myself by taking on a massive project, and I could not have done it alone. This work was designed to build upon the shoulders of giants, and it was great to have those giants on my side throughout this process. The advice you gave me from project design through to the final manuscript resulted in a better end product and far fewer headaches, which I appreciate immensely. Additionally, I would like to thank each of you for your flexibility, which allowed me to see this project through to a satisfying conclusion despite an ongoing global pandemic. As if completing a dissertation project was not hard enough!

I am especially grateful to my advisor, Mitch Sommers. I did not initially apply to complete my graduate studies as a member of the Speech and Hearing Lab, but I am so grateful that I ended up working with you. For the past five years, I have gotten to work on projects that I find truly interesting and enjoyable. Through these projects and your guidance, I have become a better researcher, critical thinker, and writer. Thank you for offering me this amazing opportunity and for preparing me for whatever adventure comes next.

A huge thank-you to three stats whizzes who helped relieve a great many headaches throughout the dissertation process. First, to Pete Millar, thank you for teaching me the basics of MPT modeling and for providing the brilliant idea to plot observed versus expected counts to visualize goodness-of-fit. Second, to Anthony Bishara, thank you for helping me troubleshoot the MPT analyses and for brainstorming theoretically-sound ways to improve model fit. Finally, to Mike

Strube, thank you for being willing to answer my many questions about mixed-effects modeling (some sent at 2am the morning before submitting the dissertation to the committee).

I would also like to thank the other members of the Speech and Hearing Lab, both past and present: to John Morton for introducing me to the topic of false hearing, which has been my passion for the last five years; to Avanti Dey for helping me over the hurdles of being a Canadian in a U.S. graduate program; to Nikki Runge for paving the way and for the helpful advice at each stage of the program; to Lauren Gaunt and Maggie Zink for sharing your eye-tracking expertise to help get my project off the ground; to Steve Dessenberger for commiserating over the many technical issues we encountered and for working with me to find timely solutions; and to Kate McClannahan for your helpful insights into academia and for the delicious rosemary sugar cookies that powered me through the second draft of the dissertation manuscript.

Also, special thanks to my many undergraduate research assistants: to Liam Gibbs for spending dozens of hours in the hot, sound-proof booth recording the stimuli for the SPIN task; and to Adam Isaacs, Aryn Lyke, Audrey Ulfers, Davis Holmes, Forrest Whiting, and Zi’Onay Walker for your hard work collecting data. Without you, this project would have taken a decade to complete!

Grad school has been difficult, but it was worth it if only for the friends I made at WashU. Thank you to the members of my band, Coefficient of Alienation – Emily Streeper, Hank Chen, Marina Gross, Ruthie Shaffer, Sam Chung, and Yu-Hua Yeh – you gave me something to look forward to at the end of each week. Thank you to Eylül Tekin, Francis Anderson, Öykü Üner, Reshma

Gouravajhala, and Sam Chung for the philosophical arguments, movie and game nights, and dinner parties that made these past five years some of the happiest of my life. I could not have asked for better people with whom to share these experiences.

Most of all, thank you to Reshma Gouravajhala. You have been my partner in all things, but your support has perhaps been most tangible throughout this dissertation process. From commenting on project ideas to aiding in hours-long searches for pictures for the visual world task, your influence is felt throughout this project. You have elevated my successes, softened my failures, and kept me sane through long writing days (and nights). I am forever grateful for your support.

Lastly, I would like to thank my parents for their unending love and support. Moving away from my home and country to start grad school was the hardest thing I have done in my life. Thank you both for believing in me and for letting me go on this adventure. It has been one of my greatest joys to get to share my accomplishments with you. Here is one more.

Eric Failes

Washington University in St. Louis

January 2021

ABSTRACT OF THE DISSERTATION

Testing the Common-Mechanisms Theory of False Hearing and False Memory: The Roles of Executive Functioning and Inhibitory Control

by

Eric Failes

Doctor of Philosophy in Psychological & Brain Sciences

Washington University in St. Louis, 2020

Mitchell S. Sommers, Chair

Recent studies have shown that older adults are more susceptible to context-based misperceptions in hearing than are younger adults, a phenomenon that has been referred to as false hearing (Rogers et al., 2012; Sommers et al., 2015). The authors of these studies have noted similarities between false hearing and false memories (Jacoby et al., 2005), suggesting that the two phenomena may arise from similar cognitive mechanisms. The present dissertation project investigated similarities between false hearing and false memories. In Experiment 1, I directly compared susceptibility to false hearing and false memories in younger and older adults. I then investigated two cognitive mechanisms that could underlie these phenomena: inhibitory control and executive functioning. I found that poorer executive functioning was related to increased susceptibility to both false hearing and false memory, supporting executive functioning as a common cognitive mechanism underlying these phenomena. In Experiment 2, I tested whether the predictive strength of sentences influenced susceptibility to false hearing, building upon the finding of a strong relationship between backwards associative strength and susceptibility to false memory (Deese, 1959; Roediger et al., 2001b). I found that the predictive strength of sentences was positively associated with susceptibility to false hearing, supporting the idea that increasing inhibitory control demands increased the likelihood that false hearing would occur. Finally, I present eye-tracking data suggesting that predictive sentences increased activation of a single word, and that older adults were less likely than younger adults to suppress this activation when a different word was presented. The findings of these experiments support the idea that false hearing and false memory share at least one common cognitive mechanism, but highlight differences between these phenomena and between younger and older adults.

Keywords: false hearing, false memory, speech perception, context effects, aging, eye-tracking

Chapter 1: Introduction

The ability to comprehend speech is one that is largely taken for granted because speech perception and comprehension seem almost automatic in one's native language. The average conversational speech rate is estimated to be between 140 and 180 words per minute (Miller et al., 1984; Wingfield et al., 1999), a quickly flowing stream of sounds that a listener must parse into individual words, match to internal representations in the mental lexicon, process for meaning, and fit within the context of what has previously been said. Add to this the extra challenges imposed by difficult listening situations, such as holding a conversation in a noisy environment, and one might question how speech can be processed in such a fluid manner.

Although speech comprehension in younger, normal-hearing, adults does seem effortless and automatic, there are a number of populations who may have increased difficulty given the perceptual demands described above. For example, older adults may be expected to have an especially difficult time comprehending speech due to age-related hearing loss (Morrell et al., 1996; Sommers et al., 2011). However, despite declines in hearing acuity, speech comprehension has been shown to be relatively stable until late in life. For example, in a cross-sectional study of adults ages 20 through 90, Sommers et al. (2011) found that listening comprehension remained stable until approximately age 65, whereas hearing acuity – especially in the high audiometric frequencies – began to decline earlier in life. Indeed, there is evidence that decline in hearing acuity can begin as early as the third decade of life (Morrell et al., 1996).

Sommers et al. (2011) suggested that one explanation for their finding of preserved listening comprehension despite declines in hearing acuity is that the passages used in their comprehension tasks provided a cohesive semantic context, which may have allowed older adults to infer what was missed due to hearing loss. There is substantial evidence showing that, relative to younger adults, older adults benefit as much or more from the addition of semantic context in speech perception (Benichov et al., 2012; Dubno et al., 2000; Hutchinson, 1989; Nittrouer & Boothroyd, 1990; Pichora-Fuller et al., 1995; Rogers et al., 2012; Sommers & Danielson, 1999; Sommers et al., 2015; Wingfield et al., 1991). For example, Wingfield et al. (1991) used a word-onset gating procedure – a task in which words were presented in 50 ms increments until their first phoneme could be correctly identified – under three different conditions: 1) where the target word was preceded by no context (e.g., *The word is* _____), 2) where the word was preceded by what the authors deemed a weakly predictive sentence (i.e., a sentence with a cloze value[1] of less than .15), or 3) where the word was preceded by what the authors deemed a highly predictive sentence (i.e., a sentence with a cloze value greater than or equal to .15). The authors found that older adults needed to hear more of the target word before they could correctly identify the first phoneme than did younger adults when the target word was preceded by either no context or a weakly predictive sentence. The authors attributed this age-related deficit in the gating paradigm to impaired hearing acuity in the older adults. However, younger and older adults identified words at similar gates when the target word was preceded by a highly predictive sentence (see Figure 1). This finding suggests that older adults may be able to compensate for deficits in hearing acuity by using available contextual cues to infer what was missed due to hearing loss or cognitive decline, which could at least partly explain the Sommers et al. (2011) finding of preserved listening comprehension despite declining hearing acuity.

---

[1] A cloze value is a measure of predictive strength and refers to the proportion of participants who complete a given sentence with a specific word. For example, a sentence with a cloze value of .15 indicates that 15% of participants completed that sentence with the same word.

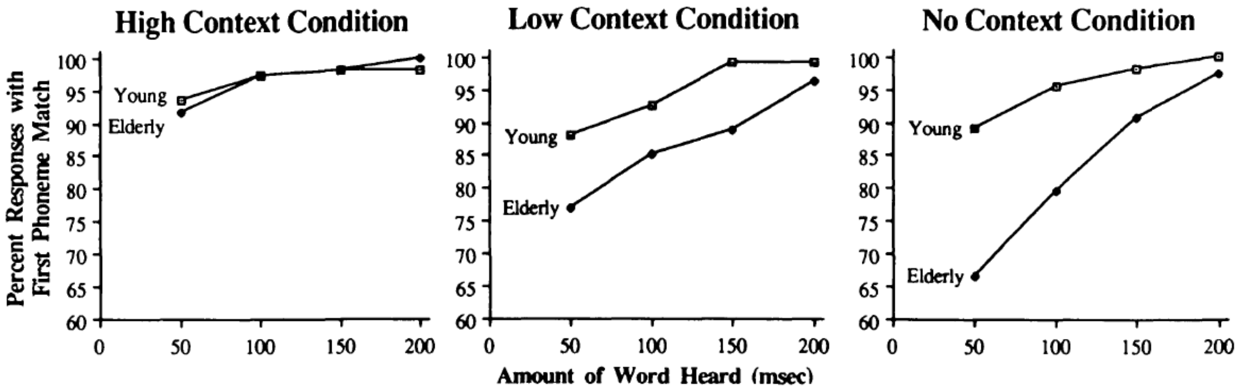

Figure 1. Figure from Wingfield et al. (1991) depicting the amount of a target word that needed to be presented before the first phoneme could be correctly identified by younger and older adults when the target word was preceded by no context, a weakly predictive sentence (low context condition), or a highly predictive sentence (high context condition).

Although semantic context can help older adults compensate for hearing loss, it can also lead older adults astray. Rogers et al. (2012), for example, demonstrated that older adults can be over-reliant on contextual cues in speech perception such that they report hearing a contextually predicted word when a similar sounding (but unpredicted) word is presented. Specifically, Rogers et al. used a process dissociation procedure (see Jacoby, 1991) to create conditions in which using context as a basis for responding would lead to a different response than if participants used the phonological content of the speech signal as a basis for responding, allowing the researchers to determine the extent to which participants relied on context. In the study by Rogers et al., this was accomplished by first having participants study semantically related cue-target word pairs (e.g., *barn – hay*) presented both visually and aurally to establish a relationship between the word pairs. After the study phase, participants completed a speech perception task in which two words were presented aurally, the first word in the clear (i.e., without background noise) and the second word in background noise. The word pairs were one of three types: 1) a cue-target pair from the study phase (congruent condition; e.g., *barn – <u>hay</u>*);

2) a cue from the study phase paired with a word differing from the learned target word by a single phoneme, known as a phonological neighbor (incongruent condition; e.g., *barn* – *pay*); or 3) a pair of unstudied, unrelated words (baseline condition; e.g., *cloud* – *fun*). Participants were tasked with identifying the word in the background noise and also reporting their confidence in the accuracy of their response from 0-100% certainty. In the congruent condition, the same response (*hay*) would be supported regardless of whether participants used context (i.e., the paired cue word) or the phonological characteristics of the target word as a basis for responding. In the incongruent condition, however, using context as a basis for responding would lead to a different response (*hay*) than using the phonological content of the target word as a basis for responding (*pay*). Thus, the authors were able to determine the extent to which participants relied on context by determining the proportion of incongruent trials in which participants erroneously provided the contextually predicted response (e.g., reporting hearing *hay* when presented with the pair *barn* – *pay*), which the authors referred to as *false hearing*.

Rogers et al. (2012, Experiment 2) found that both younger and older adults frequently experienced false hearing on incongruent trials. However, despite performance being equated across age groups in the baseline condition by manipulating the signal-to-noise ratio (SNR) individually for each participant, older adults were significantly more likely to experience false hearing than were younger adults (.39 vs. .26). Additionally, older adults experienced higher confidence in cases of false hearing than did younger adults, with older adults being almost four times more likely to report 100% confidence in cases of false hearing than younger adults (.27 vs. .07). This data showed that both younger and older adults often used context to facilitate speech perception in noise and indicated that older adults were especially reliant on contextual cues as a basis for responding.

More recent research (Sommers et al., 2015, Experiment 2) has reported that false hearing can also be evoked using sentence contexts. Instead of using learned word pairs to provide context as in Rogers et al. (2012), Sommers et al. (2015; Experiment 2) used complete sentences divided into three conditions: 1) sentences providing no context for predicting the sentence-final word (baseline condition; e.g., *He was thinking about the* *sheep*), 2) sentences providing a valid context for predicting the sentence-final word (congruent condition; e.g., *The shepherd watched his* *sheep*), and 3) sentences in which the sentence-final word was a phonological neighbor of the predicted sentence-final word (incongruent condition; e.g., *The shepherd watched his* *sheath*). In all cases, the participant’s task was to report the sentence-final word, which was presented in background noise, and to judge their confidence in the accuracy of their response on a 0-100 point scale. As in Rogers et al. (2012, Experiment 2), SNRs for the final word were set individually to obtain approximately 50% accuracy for all participants in the baseline condition. Importantly, all participants were warned that sentences would sometimes be misleading, and there were three times as many incongruent sentences as congruent sentences. Both the warning instructions and the disproportionate use of incongruent sentences should have discouraged a context-based response strategy.

The proportion of accurate responses (hits) in the congruent and baseline conditions and the proportion of context-based misperceptions in the incongruent condition (i.e., cases of false hearing) from Sommers et al. (2015, Experiment 2) are shown in panel A of Figure 2, and the confidence associated with those responses is shown in panel B. Despite explicitly warning participants that sentences could be misleading and instructing participants to respond based on what they heard and not based on context, Sommers et al. found that both younger and older adults experienced false hearing in the incongruent condition (e.g., reported hearing *sheep* when presented *The shepherd watched his* <u>*sheath*</u>). However, as in Rogers et al. (2012), older adults were more susceptible to false hearing than were younger adults (.50 vs. .39), and were more confident in cases of false hearing than were younger adults, with older adults reporting maximum confidence in cases of false hearing four times as often as younger adults (.16 vs. .04). Participants' continued use of contextual cues in conditions that discouraged use of context lends further support to the argument that both age groups – but especially older adults – relied on context as a basis for responding.

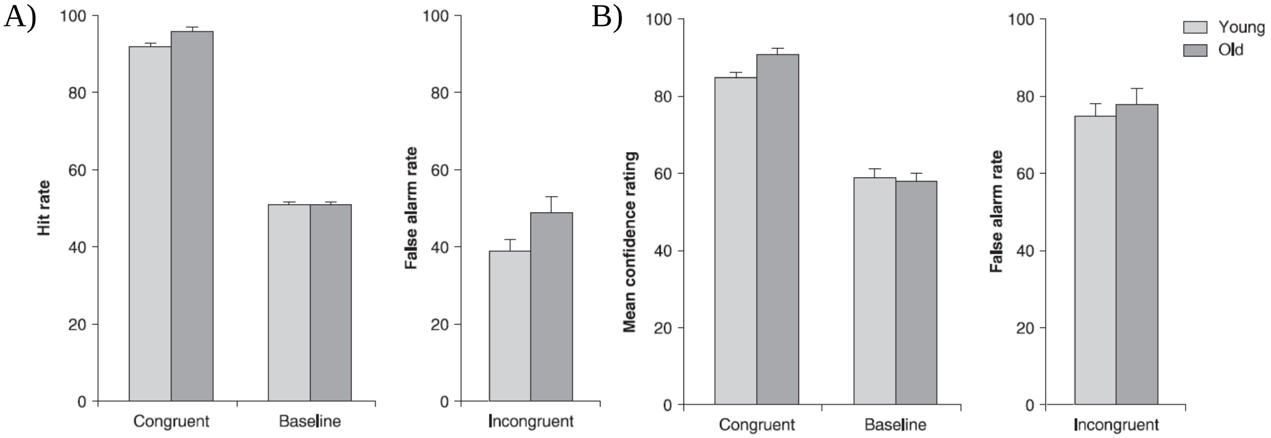

Figure 2. Plots from Sommers et al. (2015; Experiment 2) depicting the proportion of correct responses in the congruent and baseline conditions and cases of false hearing (false alarms) in the incongruent condition (A) and the confidence in congruent and baseline condition hits and cases of false hearing in the incongruent condition (B).

1.1 False memory and similarities to false hearing

As in speech perception, memory can also be biased by contextual cues and prior knowledge. Bartlett (1932) was among the first researchers to study how memory can be distorted by prior knowledge. In his seminal work, Bartlett had participants read the short story "The War of the Ghosts" and asked them to recall the story verbatim. Bartlett found that, when recalling the story, participants often changed details to better conform to their own experiences. For example, whereas the story depicts two men hunting seals, participants often falsely recalled that the two men had been fishing. Bartlett's study is often considered one of the first studies demonstrating *false memory*, an umbrella term used to describe a variety of memory distortions.

Since Bartlett's (1932) study, several paradigms have been developed to study false memory, perhaps the most prevalent being the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In the DRM paradigm, participants study lists of words in which each word in the list is semantically related (or sometimes phonologically related, see Finley et al., 2017; Sommers & Huff, 2003; Sommers & Lewis, 1999; Watson et al., 2003) to an unpresented critical word. For example, for the critical word needle, participants might study the words thread, pin, eye, sewing, etc., but would not study the word needle. Roediger and McDermott (1995, Experiment 1) found that when asked to remember the studied words, participants falsely reported having studied the critical word in 40% of cases in a recall test and in 84% of cases in a recognition test. Additionally, participants indicated that they were sure they had studied the critical word in 58% of cases in the recognition test, demonstrating high confidence in their false memories.

A number of other paradigms have also been used to study false memories. In schema paradigms like Bartlett's (1932) study described above, researchers test how participants' expectations based on past experiences influence memory. Brewer and Treyens (1981), for example, found that after being left alone in a graduate student's office for 35 seconds, participants falsely recalled seeing at least one item, on average, that might be expected to be in this setting (e.g., a book) even though the item was not present. In misinformation paradigms, participants witness an event (through films, written narratives, or lists), then are asked questions about the event before completing a final memory test. Critically, some of the post-event questions introduce details that were not present in the original event. Loftus (1975, Experiment 4) showed that nearly 30% of participants incorporated incorrect information introduced through post-event questioning into their memory for an event after a one-week delay. In pragmatic inference paradigms (Brewer, 1977; Chan & McDermott, 2006; Chan & McDermott, 2007), participants study sentences for which an inference regarding the outcome can be made, although the inferred outcome is not guaranteed. For example, when reading the sentence “The safe-cracker put the match to the fuse,” the reader may infer that the fuse was lit, but this outcome is not guaranteed. Brewer (1977) found that participants were more likely to falsely recall pragmatic inferences of a studied sentence (e.g., The safe-cracker lit the fuse) than the studied sentence verbatim (.26 vs. .19). Finally, having participants repeatedly imagine that an event occurred increases participants’ likelihood of falsely remembering that the event occurred (Goff & Roediger, 1998), a phenomenon known as imagination inflation. Thus, false memories can be evoked using diverse methods.

Interestingly, the patterns of age differences observed by Rogers et al. (2012) and Sommers et al. (2015) in false hearing are remarkably similar to those observed in past studies of false memory: Older adults have been shown to be more susceptible to false memory than younger adults in the DRM paradigm (Balota et al., 1999; Kensinger & Schacter, 1999; Watson et al., 2001) and a number of variations of the misinformation paradigm (Jacoby, 1999; Jacoby et al., 2005; Mitchell et al., 2003; Roediger & Geraci, 2007; but see Huff & Umanath, 2018; Prull & Yockelson, 2013 for equal performance across age groups). Additionally, older adults have been shown to be more confident in the accuracy of their false memories than younger adults (Jacoby et al., 2010; Mitchell et al., 2003), just as older adults were more confident in cases of false hearing than younger adults in the studies by Rogers et al. (2012) and Sommers et al. (2015).

One study of false memory that is particularly useful for comparison with studies of false hearing was conducted by Jacoby et al. (2005) who, like Rogers et al. (2012) and Sommers et al. (2015), used the process dissociation framework to separate context-based responding from, in this case, responding based on conscious recollection. As in the study by Rogers et al. described above, Jacoby et al. first had participants study semantically related word pairs (e.g., knee-bone). Participants then completed a cued fragment completion test in which they saw a cue word and a fragment of its paired target word from the study phase (e.g., knee-b_n_) and were tasked with completing the fragment with the word from the study phase (e.g., bone). Critically, before each cue-fragment pair, one of three types of primes was presented: 1) the studied target word (congruent condition; e.g., bone), 2) an unstudied word that was semantically related to the cue word and that could plausibly complete the fragment (incongruent condition; e.g., bend), or 3) a string of ampersands (baseline condition; &&&&&). These primes provided valid, invalid, and neutral cues for recalling the studied target word, respectively. Participants also characterized the subjective experience of their memory by making a remember/familiar/guess judgement. Participants were instructed to rate their response as "remembered" if they recalled specific details from when the item was presented at study, "familiar" if they knew that the word had been in the study list but could not remember specific details about its presentation, and "guessed" if they had no idea what word had been paired with the cue word in the study phase.

In line with findings from other studies of false memory (Balota et al., 1999; Jacoby, 1999; Jacoby et al., 2005; Kensinger & Schacter, 1999; Mitchell et al., 2003; Roediger & Geraci, 2007; Watson et al., 2001) and from the false hearing studies described above (Rogers et al., 2012; Sommers et al., 2015), Jacoby et al. (2005, Experiment 2) found that both younger and older adults often erroneously completed the fragment with the prime word in the incongruent condition (e.g., completing the cue-fragment pair “knee-b_n_” with “bend” when “bone” had been presented in the study phase), although older adults were more susceptible to these false memories than were younger adults (.68 vs. .48). Additionally, older adults were more than ten times as likely to characterize their false memories as “remembered” than were younger adults (.43 vs. .04), providing a conceptual replication of older adults’ elevated confidence in cases of false hearing relative to younger adults (Rogers et al., 2012; Sommers et al., 2015). These findings suggest that false memories evoked using the process dissociation framework – like cases of false hearing – result largely from overreliance on context as a basis for responding.

The similarity of their false hearing findings with those of Jacoby et al. (2005, Experiment 2) in false memory led Rogers et al. (2012) and Sommers et al. (2015) to suggest that false memory and false hearing might result, at least in part, from common cognitive mechanisms as opposed to mechanisms specific to either memory or perception. However, no published work has directly investigated the relationship between false hearing and false memory. The primary goals of the present study were to fill this gap in the literature and to investigate specific cognitive mechanisms that may underlie these two phenomena. Identifying specific mechanisms will not only advance our understanding of false hearing and false memory, but will also allow for creation of interventions to reduce susceptibility to these errors.

1.2 Potential mechanisms underlying false hearing and false memory

Although no published research has directly tested the mechanisms underlying false hearing, a wealth of research has investigated potential mechanisms underlying false memory (Bixter &

Daniel, 2013; Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Jacoby et al., 2005; Long et al., 2008; McCabe et al., 2009; Meade et al., 2012; Sommers & Huff, 2003, Experiment 2; Unsworth & Brewer, 2010; Watson et al., 2005). Several cognitive abilities have been shown to be correlated with susceptibility to false memory, but here I will focus on executive functioning and inhibitory control due to their proposed relations to false hearing in past studies (Jacoby et al., 2012; Rogers et al., 2012; Sommers et al., 2015). It is important to note that executive functioning is composed of a number of different cognitive abilities, including inhibitory control. I consider executive functioning and inhibitory control separately in the present study for two reasons. First, each of these constructs have been shown to be predictive of susceptibility to false memory (Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Jacoby et al., 2005; McCabe et al., 2009; Meade et al., 2012; Sommers & Huff, 2003, Experiment 2), and one of the primary goals of this dissertation project is to test for similarities between false hearing and false memory. Therefore, I used the same measures of executive functioning and inhibitory control that have been shown to be predictive of false memory in past studies to determine whether they were also predictive of false hearing. Second, it is important to determine whether the relationship between executive functioning and false hearing and false memory can be accounted for by inhibitory control's inclusion in the executive functioning construct. Since executive functioning is composed of multiple cognitive abilities, showing that executive functioning is related to false hearing and false memory would narrow down the list of potential cognitive abilities that could underlie these phenomena but would not single out any specific cognitive ability. Additionally, there is evidence to suggest that there are multiple functions of inhibitory control (Kramer et al., 1994; Lustig et al., 2007) – controlling what information enters working memory, removing information that is no longer relevant from working memory, and suppressing prepotent, but incorrect, responses – and only the suppression function is referenced in the inhibitory control theory of false hearing (Jacoby et al., 2012; Rogers et al., 2012; Sommers et al., 2015). Including a separate measure of the suppression aspect of inhibitory control – here, an auditory Stroop task – in addition to a measure of executive functioning allowed me to determine whether the relationship between executive functioning and false hearing and false memory persisted after statistically controlling for this aspect of inhibitory control, which would indicate that components of executive functioning aside from response suppression were related to false hearing and false memory. However, if the relationship between executive functioning and false hearing and false memory became non-significant after controlling for this aspect of inhibitory control, this would suggest that this relationship could be fully accounted for by variance attributable to differences in the ability to suppress a prepotent response.

1.2.1 Executive functioning

Baddeley and Hitch (1974) were among the first researchers to allude to executive functioning. In their model of working memory, the authors described a central executive component that oversaw the functioning of two primary storage subsystems of working memory: the phonological loop and the visuospatial sketchpad. The central executive in this model was involved in allocation of attention and improving efficiency of working memory when the capacity of the storage subsystems was exceeded by implementing encoding and retrieval strategies (e.g., chunking). Shortly thereafter, Posner and Snyder (1975) coined the term *cognitive control*, which the authors described as a limited-capacity mechanism involved in directing attention based on one’s goals. The role of cognitive control proposed by Posner and Snyder is similar to that of the modern conception of executive functioning, which is often described as the set of cognitive abilities involved in maintaining goals and directing behavior toward achieving those goals. As stated above, the modern construct of executive functioning is comprised of several cognitive abilities, and is typically thought to include inhibitory control, working memory, planning, and set shifting, although an exhaustive list of the components of executive functioning is up for debate (Jurado & Rosselli, 2007). Executive functioning is generally believed to be controlled by the frontal lobes of the brain – particularly the prefrontal cortex – as evinced by the development of executive functioning coinciding with maturation of the frontal lobes through adolescence (Anderson et al., 2001) and disorders affecting the frontal lobes (e.g., ADHD and autism) being associated with impaired executive functioning (see Jurado & Rosselli, 2007). For this reason, the term "frontal lobe functioning" is sometimes used interchangeably with executive functioning (Butler et al., 2004; Chan & McDermott, 2007; McCabe et al., 2009; McCabe et al., 2010; Meade et al., 2012; Roediger & McDaniel, 2007) and executive functioning is often assessed by measures shown to activate frontal lobe activity, such as the battery developed by Glisky et al. (1995) that was used in the present study, which will be discussed in further detail below.

Studies investigating the role of executive functioning in false memory have focused almost exclusively on explaining the increased susceptibility to false memory experienced by older, relative to younger, adults. Since older adults have been shown to be both more likely to experience deficits in executive functioning (Chan & McDermott, 2007; McCabe et al., 2009; Roediger & McDaniel, 2007) and more susceptible to false memory (Balota et al., 1999; Chan & McDermott, 2007; Jacoby, 1999; Jacoby et al., 2005; Kensinger & Schacter, 1999; Mitchell et al., 2003; Roediger & Geraci, 2007; Watson et al., 2001; but see Huff & Umanath, 2018; Prull & Yockelson, 2013 for equal performance across age groups) than younger adults, it is possible that deficits in executive functioning could increase susceptibility to false memory. Supporting this conclusion, executive functioning has been shown to partially mediate the relationship between age and susceptibility to false recognition of lures on a recognition memory test (McCabe et al., 2009). Additionally, several studies investigating the role of executive functioning in false memory have revealed that older adults' increased susceptibility to false memory relative to younger adults is partially, if not fully, eliminated when considering only older adults with high executive functioning (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007). There is also evidence that younger adults with lower executive functioning are more susceptible to false memory than those with higher executive functioning (Chan & McDermott, 2007). Therefore, it is plausible that deficits in executive functioning give rise to false memories and that older adults' increased susceptibility to false memory relative to younger adults results, at least in part, from age-related declines in executive functioning.

The authors of past studies have suggested that older adults with high executive functioning are less susceptible to false memories than those with low executive functioning due to preserved source monitoring (Butler et al., 2004; McCabe et al., 2009; Roediger & Geraci, 2007). According to this theory, better executive functioning improves one's ability to distinguish between true memory traces and information activated due to its association with available contextual cues. However, Meade et al. (2012) presented evidence to suggest that improved source monitoring cannot fully account for the relationship between executive functioning and susceptibility to false memory. The authors found that older adults with high executive functioning demonstrated decreased susceptibility to false memory of the critical word in a categorized list memory paradigm relative to older adults with low executive functioning when given a free report cued recall test, but similar performance across the high and low executive functioning groups when given a forced report cued recall test. Additionally, older adults with high executive functioning displayed equal susceptibility to false memory relative to older adults with low executive functioning when the authors controlled for guessing by subtracting the proportion of unpresented critical words that were recalled from the proportion of presented critical words that were recalled. This led the authors to suggest that differences in susceptibility to false memory between older adults with high versus low executive functioning may reflect differences in retrieval strategies. Specifically, older adults with low executive functioning may have been more likely to guess that the critical word was present due to its association with the studied list words, whereas older adults with high executive functioning may have employed a more conservative retrieval strategy. This conclusion was also endorsed by Chan and McDermott (2007), who stated that higher executive functioning may be related to increased use of controlled recollective processes, whereas individuals with lower executive functioning may rely to a greater degree on automatic, familiarity-based processing (see Jacoby, 1991; McDermott & Chan, 2006 for additional information on controlled versus automatic processing).

It is interesting to note that both Rogers et al. (2012) and Sommers et al. (2015) invoked the distinction between controlled and automatic processing to explain age differences in susceptibility to false hearing. Specifically, the authors suggested that the older adults in their studies were less able than younger adults to constrain their responding to available sensory information – which the authors called controlled or careful listening – and instead gave the automatic response that was supported by context. Sommers et al. suggested that the inability to suppress the automatic, context-based response reflected a failure of inhibitory control. Since inhibitory control is typically considered to be a component of executive functioning (see Jurado

& Rosselli, 2007), it is possible that differences in susceptibility to false memory between older adults with high and low executive functioning observed in past studies (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007) resulted from differences in inhibitory control. In the next section, I describe evidence for the role of inhibitory control in false memory and false hearing.

1.2.2 Inhibitory control

One specific component of executive functioning that has received a great deal of attention in the false memory and false hearing literatures is inhibitory control. As noted above, there are thought to be three different aspects of inhibitory control: controlling what information enters working memory, removing information that is no longer relevant from working memory, and suppressing prepotent, but incorrect, responses (Lustig et al., 2007). In support of distinct functions of inhibitory control, measures purported to tap into different functions of inhibitory control have been shown to be uncorrelated in previous studies (Kramer et al., 1994; Pineda & Merchan, 2003). For example, the number of categories achieved on the Wisconsin Card Sorting Test – a measure of the aspect of inhibitory control involved in removing no-longer-relevant information from working memory – has been shown to be uncorrelated with interference on a Stroop task (Pineda & Merchan, 2003) – a measure of the response suppression aspect of inhibitory control – and performance on the two tasks have loaded onto separate factors in exploratory factor analyses (Boone et al., 1998; Pineda & Merchan, 2003) and principle components analyses (Rodríguez-Aranda & Sundet, 2006). Thus, there are distinct aspects of inhibitory control that can be separated in different tasks.

The response suppression function of inhibitory control is often referenced in explaining false hearing (Jacoby et al., 2012; Rogers et al., 2012; Sommers et al., 2015). According to the inhibitory control theory of false hearing, misleading contextual information increases the activation of words in the mental lexicon that fit with the semantic context, thereby increasing competition for perception when the target word is unpredicted by context. For example, hearing the sentence "*The shepherd watched his*…" creates a strong expectation that the word *sheep* should follow, so the word *sheep* should become highly activated while listening to the sentence. When an unexpected but similar sounding word (e.g., *sheath*) is presented instead of the predicted word – as in the incongruent condition of the study by Sommers et al. (2015, Experiment 2) described above – participants may still perceive the predicted word if that word received more activation from context than the target word received from the auditory signal. To improve the chances of accurately hearing the unpredicted target word, it is thought that inhibitory control can be used to suppress the activation of the unpresented, but contextually predicted, word. Thus, according to this theory, false hearing occurs when people are unable to sufficiently inhibit the contextually predicted word, and older, relative to younger, adults are especially prone to false hearing due to an age-related deficit in inhibitory control (Cohn et al., 1984; Hasher & Zacks, 1988; Hasher et al., 1997; Jacoby et al., 2005; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West & Alain, 2000).

Past studies have provided at least indirect evidence to support inhibitory control as a mechanism underlying false hearing. Sommers and Danielson (1999, Experiment 2), for example, conducted a study to investigate how inhibitory control influenced veridical speech perception of lexically easy words (i.e., words with few phonological neighbors, that have relatively low frequency of occurrence in language) and lexically hard words (i.e., words with many phonological neighbors, that have relatively high frequency of occurrence in language). According to the Neighborhood Activation Model (NAM; Luce & Pisoni, 1998), hearing a word activates both the spoken word and its phonological neighbors in the mental lexicon. The NAM postulates that relative activation of these words is determined by two primary factors: 1) the phonological similarity of the activated words to the presented word, and 2) the frequency with which the activated words appear in language. Thus, lexically hard words should be harder to identify in noise than lexically easy words because they have a greater number of high-frequency phonological competitors. Sommers and Danielson suggested that inhibitory control might be needed to reduce the activation of phonological neighbors to accurately hear the target word. Consequently, they hypothesized that lexically hard words, which have more high-frequency phonological neighbors to compete for perception, should require greater inhibitory control than lexically easy words. Additionally, they hypothesized that valid contextual cues may aid speech perception by selectively increasing the activation of the contextually predicted target word in the mental lexicon, thereby decreasing the need to inhibit phonological neighbors that did not fit with semantic context.

Sommers and Danielson (1999) tested these predictions by having younger and older adults listen to sentences with the final word in noise for identification, similar to the design of the false hearing study by Sommers et al. (2015, Experiment 2). Unlike the study by Sommers et al., however, the sentences used in Sommers and Danielson’s study were never misleading. Instead, the sentences used by Sommers and Danielson provided either little context for predicting the target word (low-predictability sentences; e.g., *She was thinking about the* <u>*path*</u>) or a highly predictive context supporting the target word (high-predictability sentences; e.g., *She was walking along the* <u>*path*</u>). Participants also completed two measures of the response suppression aspect of inhibitory control – a selective attention paradigm developed by Garner

(1974) and an auditory Stroop task (Stroop, 1935) – scores from which were combined to form a composite measure of inhibitory control.

In line with their hypotheses, Sommers and Danielson (1999, Experiment 2) found that their measure of the response suppression aspect of inhibitory control was significantly correlated with identification of lexically hard, but not lexically easy, words, such that individuals with poorer inhibitory control had more difficulty identifying lexically hard words than individuals with better inhibitory control. Supporting the idea that valid contextual cues decreased the need to inhibit activated phonological neighbors of the target word, the correlation between their measure of inhibitory control and identification of lexically hard words was stronger when the word was preceded by a low-predictability sentence (r = -.73) than a high-predictability sentence (r = -.58). Together, these findings suggest that inhibitory control was needed to suppress activation of phonological neighbors of a target word, and that the availability of predictive context partially alleviated reliance on inhibitory control to achieve accurate perception.

Sommers and Danielson's (1999, Experiment 2) study established the relationship between the response suppression aspect of inhibitory control and veridical hearing but does not tell us whether inhibitory control contributes to false hearing. However, the proposal that inhibitory control is needed to reduce activation of a target word's phonological neighbors and that context increases the activation of semantically viable words is highly pertinent to the studies of false hearing discussed above. Recall that the target words in the incongruent condition of the tasks used by Rogers et al. (2012) and Sommers et al. (2015) were phonological neighbors of the word predicted by context. Based on the NAM (Luce & Pisoni, 1998) and the findings of Sommers and Danielson, we would expect that hearing the incongruent target word (e.g., *sheath* in the sentence *The shepherd watched his* *sheath*) would activate the contextually predicted phonological neighbor of the target word (e.g., *sheep*), and that the activation of this phonological neighbor would be boosted due to its compatibility with available contextual cues. This set of conditions should result in an increased probability that participants would mistakenly "hear" the contextually predicted phonological neighbor, a case of false hearing. Therefore, if the role of inhibitory control is to decrease activation of phonological neighbors of the target word, as suggested by Sommers and Danielson, then we might expect that individuals with better inhibitory control would be less susceptible to false hearing than those with poorer inhibitory control.

One experimental method that may be particularly useful for testing the inhibitory control account of false hearing described above is eye-tracking. Eye-tracking has been increasingly used to study language processing because it allows the researcher to observe changes in attentional focus over time, providing a real-time assessment of the processing that occurs before a response is made. In speech perception research, eye-tracking is often used within a visual world paradigm (Allopenna et al., 1998; Dahan & Gaskell, 2007; Ito et al., 2018; Kukona, 2020; Mishra & Singh, 2013; Revill & Spieler, 2012), in which different response options are depicted in the form of written words or pictures. It has been argued that changes in the proportion of fixations on the images in the visual world paradigm can be used as a proxy for changes in activation, with more highly activated response options receiving more attention than less highly activated options (Tanenhaus et al., 2000). Supporting this claim, images depicting high-frequency words, which are assumed to gain more activation than low-frequency words by speech perception models such as the NAM (Luce & Pisoni, 1998), tend to receive a greater proportion of fixations than images depicting low-frequency words (Dahan & Gaskell, 2007;

Dahan et al., 2001; Revill & Spieler, 2012). Similarly, following from Sommers and Danielson's (1999) claim that words gain activation when supported by contextual cues, images depicting words that are congruent with available semantic context tend to receive a greater proportion of fixations than images depicting words that do not fit with context (Ito et al., 2018; Kukona, 2020). Therefore, the eye-tracking methodology could allow for a test of the inhibitory control theory of false hearing. Misleading sentences, as in the incongruent condition of the study by Sommers et al. (2015), should lead to increased fixations on an image depicting the contextually predicted (but incorrect) word. If participants are able to suppress the activation of the contextually predicted word when the unpredicted target word is presented, fixations on the image depicting the contextually predicted word should decrease and fixations on the image depicting the target word should increase following presentation of the target word. In Experiment 2 of the present study, I used eye-tracking to test this hypothesis.

Whereas the relationship between inhibitory control and false hearing has not been tested directly in past studies, there has been empirical support for a relationship between inhibitory control and false memory (Colombel et al., 2016; Lövdén, 2003; Sommers & Huff, 2003, Experiment 2). Similar to the inhibitory control theory of false hearing, the inhibitory control theory of false memory suggests that false memories occur when an individual is unable to suppress activation of unpresented information. Using the DRM paradigm as an example, the inhibitory control theory suggests that reading a list of words in which each word is semantically (or phonologically) related to an unpresented critical word increases the activation of the critical word. False memories occur when the individual is unable to suppress activation of the critical word using inhibitory control (see Balota et al., 1999), leading to the experience of remembering the critical word even though it was not presented. In support of this theory, Sommers and Huff

(2003, Experiment 2) found that reaction time on incongruent trials of an auditory Stroop task (Stroop, 1935) – an indicator of the response suppression aspect of inhibitory control – was positively related to both false recall and false recognition in phonologically-associated DRM lists, accounting for 11% of variance in false recall and 7% of variance in false recognition after controlling for differences in age and processing speed (assessed by reaction time on baseline trials of the auditory Stroop task). Similarly, Colombel et al. (2016) found that a composite of four measures of inhibitory control predicted false recall of critical words in the DRM paradigm, also accounting for 11% of the variance in susceptibility to false recall. Thus, poor inhibitory control has been shown to be related to increased susceptibility to false memory, at least within the DRM paradigm.

Further evidence for the role of inhibitory control in false memory comes from the capture model (see Figure 3), a multinomial processing tree (MPT) model that has been shown to explain age differences in susceptibility to false memories evoked using the process dissociation framework (Jacoby et al., 2005, Experiment 2). The capture model used by Jacoby et al. (2005, Experiment 2; see study description above) had five parameters: 1) a capture parameter (C) corresponding to the probability of being captured by available contextual cues, in which case participants respond with the prime without attempting recollection; 2) a recollection parameter (R) corresponding to the probability of successfully remembering the studied target word when the individual was not captured by context; 3) an accessibility bias parameter (A) reflecting the probability of responding with the easily accessible prime following an unsuccessful attempt at recollection; 4) a word generation parameter (W) corresponding to the probability of generating the pre-determined alternative response on a given trial; and 5) an attribution threshold parameter (AT) signifying the likelihood of characterizing a response as “remembered” or “familiar” versus “guessed.” The authors suggested that the capture parameter represented inhibitory control, indexing the ability to resist the context-based response. They were then able to test the contribution of inhibitory control to responding in their memory task by comparing the fit of an MPT model that did not include the capture parameter to a model that included the capture parameter.

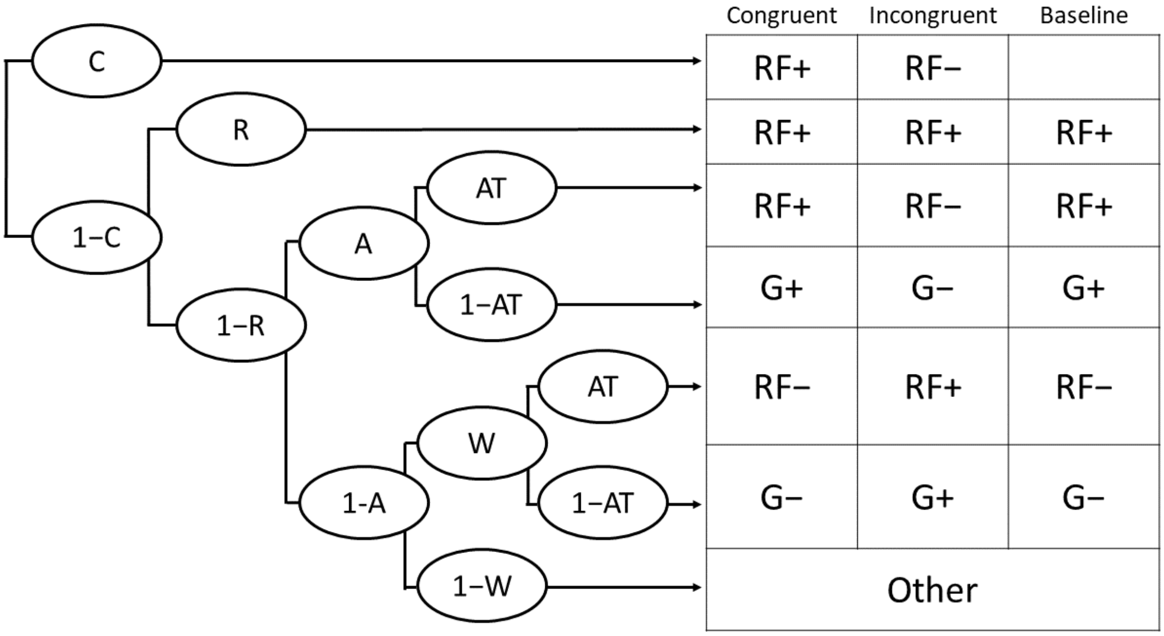

Figure 3. The capture model from Jacoby et al. (2005, Experiment 2). The table indicates performance in the three prime conditions (congruent/incongruent/baseline). Correct responses are signified by a plus sign (+), incorrect responses are signified by a minus sign (−), and “Other” refers to responses other than the target word or the pre-determined alternative word. Responses are characterized as either “remembered” or “familiar” (RF) or as “guessed” (G).

Supporting the role of inhibitory control in their memory task, Jacoby et al. (2005, Experiment 2) found that model fit was improved by adding the capture parameter. This finding was important, as it suggested that an additional means of achieving an accurate response in the congruent condition and a false memory response in the incongruent condition other than remembering the target word (the recollection parameter), responding based on ease of accessibility in memory (the accessibility bias parameter), or guessing (the word generation parameter) was necessary to fit the data. This provided evidence that capture – interpreted as representing inhibitory control – contributed to accurate responses in the congruent condition and false memory responses in the incongruent condition in the memory task.

In addition to finding that capture influenced responding in the memory task, the MPT model created by Jacoby et al. (2005, Experiment 2) also provided evidence that age differences in their memory task could be fully accounted for by differences in capture. When allowing all of the parameters in their MPT model to vary across age groups, the authors found that only the capture parameter differed significantly between younger and older adults. In fact, the model estimated that younger adults were never captured by context (capture parameter = .00) whereas older adults had a 38% chance of being captured by context. This aligns with previous studies suggesting that inhibitory control is better in younger adults than older adults (Cohn et al., 1984; Hasher & Zacks, 1988; Hasher et al., 1997; Jacoby et al., 2005; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West & Alain, 2000), and suggests that younger adults' false memories may have resulted predominantly from accessibility bias (i.e., the other parameter in the MPT model that gives rise to false memory responses) whereas older adults' false memories may have reflected a combination of accessibility bias and failures of inhibitory control. Therefore, it is possible that differences in inhibitory control can fully account for older adults' increased susceptibility to false memory relative to younger adults, and that different mechanisms may give rise to false memories in younger and older adults.

The ability of the capture model to account for age differences in false memory may be important for advancing our understanding of the mechanisms underlying false hearing. As noted above, the methods and results of the study of false memory by Jacoby et al. (2005, Experiment 2) were quite similar to those of the study of false hearing by Rogers et al. (2012). Additionally, much as the false memories in the Jacoby et al. study were described as resulting from being captured by an easily accessible prime, the words that were falsely heard in the studies by Rogers et al. and Sommers et al. (2015) were also easily accessible due to their association with available contextual cues. Therefore, it is possible that both false memory and false hearing arise from being captured by available contextual cues (i.e., failures to inhibit the contextually predicted response). Together, these similarities suggest that the capture model might be a good fit to false hearing data.

It is important to note that although Jacoby et al. (2005) interpreted the capture parameter as representing inhibitory control, this is not necessarily the case. The authors' interpretation of capture was based on the idea that capture occurs when participants are unable to resist the prepotent response – resulting in accurate responses in the congruent condition and false memory responses in the incongruent condition – a description that aligns with that of the response suppression aspect of inhibitory control (Lustig et al., 2007). However, the capture model does not manipulate any factors to isolate cognitive processes, but instead estimates parameters based on the patterns of responses across context conditions that would be expected if responding was influenced by a particular cognitive ability. Thus, it is possible that the capture parameter could reflect processes other than inhibitory control that would result in a similar pattern of responding. For example, the capture parameter could reflect personality factors such as conscientiousness, with less conscientious participants being less likely to follow instructions or pay attention to the target word, instead using context as a basis for responding to decrease listening effort. Whereas personality factors such as this may be partially accounted for by other parameters in the model, such as the accessibility bias parameter, these factors could influence the capture parameter. Therefore, finding that adding the capture parameter improves model fit is weak evidence in isolation for the role of inhibitory control in false hearing and false memory. In conjunction with other evidence supporting the role of inhibitory control, however, finding that adding the capture parameter improves model fit may increase our confidence that inhibitory control influences responding.

1.3 The present study

In Experiment 1 of the present study, I examined whether the age-related increase in susceptibility to false hearing observed in past studies (Rogers et al., 2012; Sommers et al., 2015) could be accounted for by two related mechanisms that have been shown to explain age differences in susceptibility to false memory: executive functioning and the response suppression aspect of inhibitory control. I used a speech perception in noise (SPIN) task that was nearly identical to that used by Sommers et al. (2015, Experiment 2) in which participants identified words in noise at the end of three types of sentences: 1) sentences providing no context (baseline condition; e.g., *The word is* <u>*page*</u>), 2) sentences providing a valid semantic context (congruent condition; e.g., *She put the toys in the* <u>*box*</u>), or 3) sentences providing a misleading semantic context (incongruent condition; e.g., *She put the toys in the* <u>*fox*</u>). Participants characterized their responses as either “heard,” “known,” or “guessed” to correspond to the remember/familiar/guess judgements made in the study of false memory by Jacoby et al. (2005, Experiment 2). Participants also completed the memory task used by Jacoby et al., allowing for direct comparison of susceptibility to false hearing and false memory.

In addition to the SPIN and memory tasks, participants completed a battery of executive functioning measures (Glisky et al., 1995) and an auditory Stroop task (Stroop, 1935) – each of which has demonstrated significant relationships with false memory (Butler et al., 2004; Chan & McDermott, 2007; McCabe et al., 2009; Meade et al., 2012; Roediger & Geraci, 2007; Sommers & Huff, 2003, Experiment 2) – to assess the relationships between executive functioning, inhibitory control, false hearing, and false memory. To further assess the role of inhibitory control in false hearing and false memory, the capture model created by Jacoby et al. (2005, Experiment 2) was fit to the data from the SPIN and memory tasks to determine 1) whether adding a capture parameter improved model fit relative to models that did not account for capture, and 2) the extent to which differences in responding between younger and older adults could be explained by differences in capture. To fit the capture model to data from the SPIN task, the recollection parameter from the original capture model was replaced with a hearing parameter, which corresponded to the likelihood of accurately hearing the target word in noise.

This experiment provided the first direct test of the relationship between false hearing and false memory, and was the first to examine the roles of executive functioning and inhibitory control in false hearing. Additionally, this was the first use of the capture model to investigate the cognitive processes contributing to false hearing. Consequently, the findings of this experiment will have important implications for our understanding of speech perception and memory, separately, and for formulation of a united theory of context-based errors across perceptual and memory systems.

Chapter 2: Experiment 1

2.1 Methods

2.1.1 Participants

An a priori power analysis was conducted in G*Power 3.1 (Faul et al., 2009) to determine the sample size needed to detect a correlation of -.32 – the correlation between executive functioning and false alarms characterized as "remembered" on a recognition memory test (i.e., false memories) from McCabe et al., (2009), who used the same executive functioning battery as in the present study (Glisky et al., 1995) – with a power of .80. The power analysis revealed that 56 younger and 56 older adults would be needed. Based on the results of this power analysis, data was collected from 60 younger adults ages 18-23 (*M* = 19.05, *SD* = 1.05) and 58 older adults ages 65-81 (*M* = 71.17, *SD* = 4.14) for Experiment 1.

All participants were native English speakers who did not require the use of a hearing aid and who self-reported normal or corrected-to-normal vision. Participants completed a demographic questionnaire gathering information about gender, age, handedness, years of education, ethnic group, occupation or area of study, domestic arrangement (i.e., live alone or with others), current/past health issues that could affect performance, and number of languages spoken in addition to English.

2.1.2 Hearing acuity

Hearing thresholds were assessed for octave frequencies from 250 to 8000 Hz in a sound-attenuating booth using standard pure-tone audiometry. Best-ear pure-tone averages (PTA) across the 500, 1000, and 2000 Hz frequencies[2] differed significantly between younger (*M* = 4.72, *SD* = 4.51) and older adults (*M* = 17.39, *SD* = 9.74), *t*(79.78) = -9.01, *p* < .001. Therefore, the older adults in this study exhibited poorer hearing acuity, on average, than the younger adults, as would be expected due to age-related hearing loss (Morrell et al., 1996; Sommers et al., 2011).

2.1.3 Speech Perception in Noise (SPIN) task

2.1.3.1 SPIN task stimuli

Stimuli in the SPIN task were 42 carrier sentences (baseline sentences; e.g., *The word is* <u>*page*</u>) and 42 high-predictability sentences (congruent sentences; e.g., *She put the toys in a* <u>*box*</u>) selected from the Revised Speech Perception in Noise Test (Bilger et al., 1984) or created specifically for this study. For each congruent sentence, an incongruent sentence was created by changing either the first or last sound in the sentence-final word (first sound changed in 85% of cases) to form an alternative word that was not predicted by the sentence context (e.g., *She put the toys in the* <u>*fox*</u>). The length, frequency, and phonological neighborhood density of all target words were gathered from the English Lexicon Project (Balota et al., 2007). The length of target words in the congruent (*M* = 4.35, *SD* = .99), incongruent (*M* = 4.36, *SD* = 1.05), and baseline conditions (*M* = 4.08, *SD* = .73) did not differ significantly, *F*(2,197) = 1.35, *p* > .05. The frequency of target words in the congruent (*M* = 53,839, *SD* = 120,601.25, incongruent (*M* = 57,573, *SD* = 86,165.11), and baseline conditions (*M* = 59,267, *SD* = 115,873.36) also did not differ significantly, *F*(2,197) = .04, *p* > .05. Finally, the phonological neighborhood density of target words in the congruent (*M* = 19.22, *SD* = 11.03), incongruent (*M* = 19.20, *SD* = 11.45), and baseline conditions (*M* = 18.95, *SD* = 9.73) did not differ significantly, *F*(2,197) = .01, *p* >

---

[2] The 500, 1000, and 2000 Hz frequencies were chosen because these frequencies are known to be important for speech perception (Dobie, 2011).

.05. Congruent and incongruent sentences were counterbalanced such that the congruent and incongruent versions of a sentence appeared an equal number of times across participants but with only one version appearing in a single experimental session. All sentences were recorded at 44,100 Hz and 16-bit resolution in a double-walled, sound-attenuating booth, and were spoken at a normal rate by a male with a Midwestern American accent. All sentences were played at an average amplitude of 64 dB sound pressure level (SPL).

2.1.3.2 SPIN task procedure

Baseline, congruent, and incongruent sentences were played through headphones with the final word presented in speech-shaped noise for identification. The SNR was set to -4 dB SPL for younger adults and to +1 dB SPL for older adults, SNRs shown to achieve approximately 50% target word identification accuracy in the baseline condition in pilot testing. Participants completed six practice trials divided equally among baseline, congruent, and incongruent stimuli. They then completed 120 test trials composed of 40 baseline trials, 40 congruent trials, and 40 incongruent trials. Stimuli in each counterbalancing condition were randomized once and that order was used for all participants in that condition. Participants were instructed to identify the word in noise, and to type their response after hearing each sentence. Participants were warned that sentence contexts could be misleading, and were given an example of an incongruent sentence.

After typing their response, participant completed a version of the remember/know paradigm (Tulving, 1985) – a hear/know/guess paradigm – in which they were instructed to indicate whether they heard the word they typed on the previous screen, did not hear the word they typed on the previous screen but knew that it was the word that was presented, or simply guessed the word they typed on the previous screen. Descriptions of the hear, know, and guess responses were presented on screen each time participants made this judgement. Participants indicated their response by pressing the number key corresponding to their judgement. This task was self-paced, with participants having an unlimited amount of time to type their response for the SPIN task and to make their hear/know/guess judgement.

It should be noted that context was only a valid predictor of the target word in one-third of sentences. This, in addition to the warning participants received regarding the presence of incongruent sentences, should have biased participants against using the sentence context as a basis for responding.

2.1.4 Memory task

2.1.4.1 Memory task stimuli

Susceptibility to false memory was assessed using the memory task from Jacoby et al. (2005, Experiment 2). Participants completed two phases in this task, a study phase and a test phase. Stimuli in the study phase were 99 semantically related cue-target word pairs (e.g., head – skull). Cue words ranged from three to nine letters in length (*M* = 5.08, *SD* = 1.41), and target words ranged from three to seven letters in length (*M* = 4.59, *SD* = .79). Stimuli in the test phase consisted of the same cue words from the study phase matched with a fragment of its paired target word created by replacing certain letters with dashes (e.g., head – s--l-). Target words in the test phase had between one and five letters replaced by dashes (*M* = 2.21, *SD* = .71), and retained between one and four letters (*M* = 2.38, *SD* = .63), although all but one fragment (-u- for the target word “mug”) retained at least two letters. Prior to presenting the target fragment in the test phase, participants saw one of three types of prime words: 1) the target word from the study phase (congruent condition; e.g., skull), 2) an alternative word that could complete the target fragment and that was semantically related to the cue word (incongruent condition; e.g., scalp), or 3) a string of five ampersands (&&&&&), which provided no predictive value for the fragment completion task (baseline condition). There were no differences in target word length between the congruent (*M* = 4.57, *SD* = .73), incongruent (*M* = 4.47, *SD* = .78), and baseline conditions (*M* = 4.73, *SD* = .87), *F*(2,87) = .87, *p* > .05. I obtained the log frequencies for target words in the congruent (*M* = 9.37, *SD* = 2.01), incongruent (*M* = 9.67, *SD* = 1.55), and baseline conditions (*M* = 9.29, *SD* = 1.74) from the English Lexicon Project (Balota et al., 2007), and found that there were no differences in target word frequency across conditions, *F*(2,87) = .38, *p* > .05.

2.1.4.2 Memory task procedure

Participants were told to study a list of semantically related word pairs for a later memory test. They were shown six cue-target word pairs as examples. Following the examples, participants completed the study phase, in which 90 cue-target word pairs were presented. An additional three cue-target word pairs were presented following the study phase to act as a buffer against the recency effect. For younger adults, each word pair was presented for 1000 ms, whereas each word pair was presented for 3000 ms for older adults, a manipulation used by Jacoby et al. (2005, Experiment 2) to equate memory for words in the baseline condition. An interstimulus interval (ISI) of 500 ms separated each cue-target word pair. To equate the interval from the first word in the study phase to the last word in the buffer trials, younger adults were shown a screen following the buffer trials for 186 seconds instructing them to relax and wait for the memory test to begin in three minutes, whereas older adults started the memory test immediately following the buffer trials. Stimuli in the study phase were presented in the same pre-determined random order for all participants, and each cue-target word pair was presented only once.

In the memory test, participants were again shown the word pairs from the learning phase, but in each case the second word in the pair was fragmented by substituting dashes for some of the letters. Participants were instructed to complete the word fragment by typing the full word that accurately completed the learned word pair from the study phase. As noted, before each cue-fragment pair was presented, a congruent, incongruent, or baseline prime was flashed on screen. Participants were explicitly warned at the start of the study that primes would appear, and were told that the prime would either be the target word from the study phase, in which case they should use this word to complete the word fragment, an unstudied word that was semantically related to the presented cue and that could complete the word fragment, in which case they should not use this word to complete the word fragment, or a string of ampersands. The prime remained on screen for 500 ms and offset 500 ms before presentation of the cue-fragment pair. Participants completed 30 trials in each of the three prime conditions (baseline, congruent, and incongruent), and trials were presented in a pre-determined random order.

After typing the word that completed each cue-target word pair, participants indicated their reason for providing their response by selecting one of three options: remember, know, or guess. Participants were told to select “remember” if they recalled specific details of the word’s presentation in the study list, so they were certain that their response was the word from the study list. They were told to select “know” if they knew that they had studied the word they typed, but could not recall specific details about studying it. Finally, they were told to select “guess” if they were purely guessing with no idea of the correct answer. The descriptions for each these response types were presented on screen each time participants made the remember/know/guess judgement. Participants indicated their response by pressing the number key corresponding to their judgement. This task was self-paced, with participants having an unlimited amount of time to type their response for the fragment completion task and to make their remember/know/guess judgement.

As was true of contextual cues in the SPIN task, the prime was only an accurate cue for memory in one-third of trials, and participants were warned that the prime could be misleading. Therefore, if anything, participants should have been biased against using the prime as a basis for responding.

2.1.5 Shipley Vocabulary Test

Vocabulary knowledge was assessed using the Shipley Vocabulary Test (Shipley, 1940). Participants completed 40 trials in which they decided which of four words was most similar in meaning to a target word, and indicated their responses by pressing the number key corresponding to their answer. The target word was presented at the top of the screen in capital letters, and the four numbered response options were presented horizontally below. Participants had an unlimited time to select their response. An interval of 1000 ms separated the input of a response and the onset of the next trial. Older adults (M = 34.98, SD = 2.98) displayed significantly better vocabulary knowledge than younger adults (M = 30.83, SD = 3.61), $t(112.80) = -6.79$, $p < .001$.

2.1.6 Auditory Stroop task

Individual differences in the response suppression aspect of inhibitory control were assessed using the auditory Stroop task (Stroop, 1935) from Sommers and Huff (2003, Experiment 2). Stimuli in this task were three spoken words (*male*, *female*, and *person*) spoken by four different talkers (two male and two female). Combinations of the words and talkers created three conditions: 1) a neutral condition in which the word *person* was spoken by either a male or female talker, 2) a congruent condition in which the word *male* or *female* was spoken by a talker of the same gender (e.g., a male talker saying *male*), and 3) an incongruent condition in which the word *male* or *female* was spoken by a talker of the opposite gender (e.g., a female talker saying *male*). Participants were tasked with pressing the "z" key on the keyboard if the speaker was male and the "/" key if the speaker was female as quickly and accurately as possible, and their response times and accuracy were recorded. Participants completed 24 trials in each condition for a total of 72 test trials. As in the study by Sommers and Huff, inhibitory control was assessed by the response time in the incongruent condition after controlling for response time in the baseline condition, indexing the increased processing time needed to suppress responding with the spoken word to accurately respond with the gender of the speaker.

Prior to completing the test trials, participants heard each sound stimulus and were tasked with saying aloud the gender of the speaker and the word they said to ensure that participants could accurately identify the gender of the speaker and the spoken word. Participants accurately identified 98.93% of stimuli, on average. Participants then completed 12 practice trials divided equally between baseline, congruent, and incongruent conditions before beginning test trials. All auditory stimuli were recorded at 44,100 Hz and 16-bit resolution in a double-walled, sound-attenuating booth, and were played at an average amplitude of 64 dB SPL.

2.1.7 Executive functioning battery

Executive functioning was assessed using a battery of five measures identified by Glisky et al. (1995) as tapping into frontal lobe functioning – often used interchangeably with executive functioning (see McCabe et al., 2010) – that has since been used in several studies of false memory (Butler et al., 2004; Chan & McDermott, 2007; McCabe et al., 2009; Meade et al., 2012; Roediger & Geraci, 2007). The tests were a computerized version of the modified Wisconsin Card Sorting Test (Hart et al., 1988), the Arithmetic test from the Wechsler Adult

Intelligence Scale – Revised (WAIS-R; Wechsler, 1981), the Mental Control test and Backward Digit Span from the Wechsler Memory Scale – III (WMS-III; Wechsler, 1997), and the Controlled Oral Word Association Test (Spreen & Benton, 1977). Following the procedure used by Chan and McDermott (2007), the outcome measures from each of these tests were standardized in the full sample then averaged together to form a composite executive functioning score for each participant.

2.1.7.1 Wisconsin Card Sorting Test

In the modified Wisconsin Card Sorting Test (Hart et al., 1988), participants saw a series of target cards one at a time and sorted them into one of four piles based on a changing sorting rule. Cards could be sorted based on the shape depicted on the card, the number of shapes on the card, or the color of the shapes on the card (the fourth pile did not match the target card on any of these features). Participants were instructed to click on the pile that they believed the target card belonged to based on the current sorting rule. Participants were not told the current sorting rule but were given feedback after each trial indicating whether or not they had sorted the card correctly. The sorting rule changed after the participant successfully sorted six cards in the current response category. As in the study by Hart et al. (1988), the order of categories was color, shape, number, color, shape, number. Participants continued until they had completed all six response categories or until 72 trials elapsed. The number of categories completed within the allotted 72 trials was the outcome measure.

2.1.7.2 WAIS-R Arithmetic test

In the Arithmetic test from the WAIS-R (Wechsler, 1981), the experimenter read arithmetic problems aloud and the participant was tasked with solving the problems as quickly as possible while being accurate. Participants were not allowed to write anything down, but the experimenter would repeat the question once if requested. Participants were scored based on the accuracy of their response and received bonus points on questions 10-14 if the correct answer was given in an amount of time specified on the WAIS-R scoring packet. Response time was recorded using a handheld stopwatch.

2.1.7.3 WMS-III Mental Control test

In the Mental Control test from the WMS-III (Wechsler, 1997), participants were instructed to say a sequence (e.g., list the days of the week in forward order starting with Sunday) aloud in the proper order as quickly as possible. Participants were scored based on the number of errors committed and received bonus points on perfectly accurate responses based on their response time in accordance with the WMS-III scoring packet. Response time was recorded using a handheld stopwatch.

2.1.7.4 WMS-III Backward Digit Span

In the Backward Digit Span from the WMS-III (Wechsler, 1997), the experimenter said strings of numbers of increasing length aloud and participants were instructed to repeat the numbers aloud in reverse order. Participants completed two strings of lengths two through eight and testing was stopped either after all strings were completed or when participants were unable to perfectly recall all numbers in both trials at a given string length. Participants were told that there was no time limit, and received one point for each string recalled perfectly.

2.1.7.5 Controlled Oral Word Association Test

In the Controlled Oral Word Association Test (Spreen & Benton, 1977), participants said aloud as many words as they could think of (other than proper names) starting with the letters *F*, *A*, and *S*. The outcome measure was the total number of distinct words generated across the three letters given one minute for each letter. Responses were recorded by hand by the experimenter, and a one-minute timer was used to indicate when the allotted time had elapsed.

2.1.8 General procedure

The experiment took place across two one-hour sessions on different days. During the first session, participants read a consent document and completed the demographics form, then completed the hearing test, SPIN task, auditory Stroop task, and Shipley Vocabulary Test. During the second session, participants completed the memory task, Wisconsin Card Sorting Test, Arithmetic test, Mental Control test, Backward Digit Span, and Controlled Oral Word Association Test, in that order. Participants were offered breaks between each of the tasks. Student participants received one course credit and older adults received a total of $15 ($10 + $5 to cover transportation costs) after completing each session of the experiment.

2.1.9 MPT modeling procedure

The capture model proposed by Jacoby et al. (2005, Experiment 2) – an MPT model – was fit to the data from the SPIN and memory tasks using the fit.mpt function from the *MPTinR* package in R (Singmann & Kellen, 2013). The model fit to the memory task data was the same as used by Jacoby et al. (2005, Experiment 2; see Figure 3). The model fit to the SPIN task data can be seen in Figure 4. Similar to the capture model proposed by Jacoby et al. (2005, Experiment 2), the model fit to the SPIN data had five parameters: 1) a capture parameter (C) representing the likelihood of responding based on context without listening to the target word, 2) a hearing parameter (H) representing the likelihood of accurately hearing the target word when capture did not occur, 3) an accessibility bias parameter (A) representing the likelihood of responding with the most easily accessed response (i.e., the contextually predicted word in the congruent/incongruent conditions and the presented word in the baseline condition) following an unsuccessful attempt to hear the target word, 4) a word generation parameter (W) representing the likelihood of providing the pre-designated alternative word (e.g., the alternative word for *box* in the sentence *She put the toys in the* <u>*box*</u> is *fox*), and 5) an attribution threshold parameter (AT) signifying the likelihood of characterizing a response as “heard” or “known” versus “guessed.”

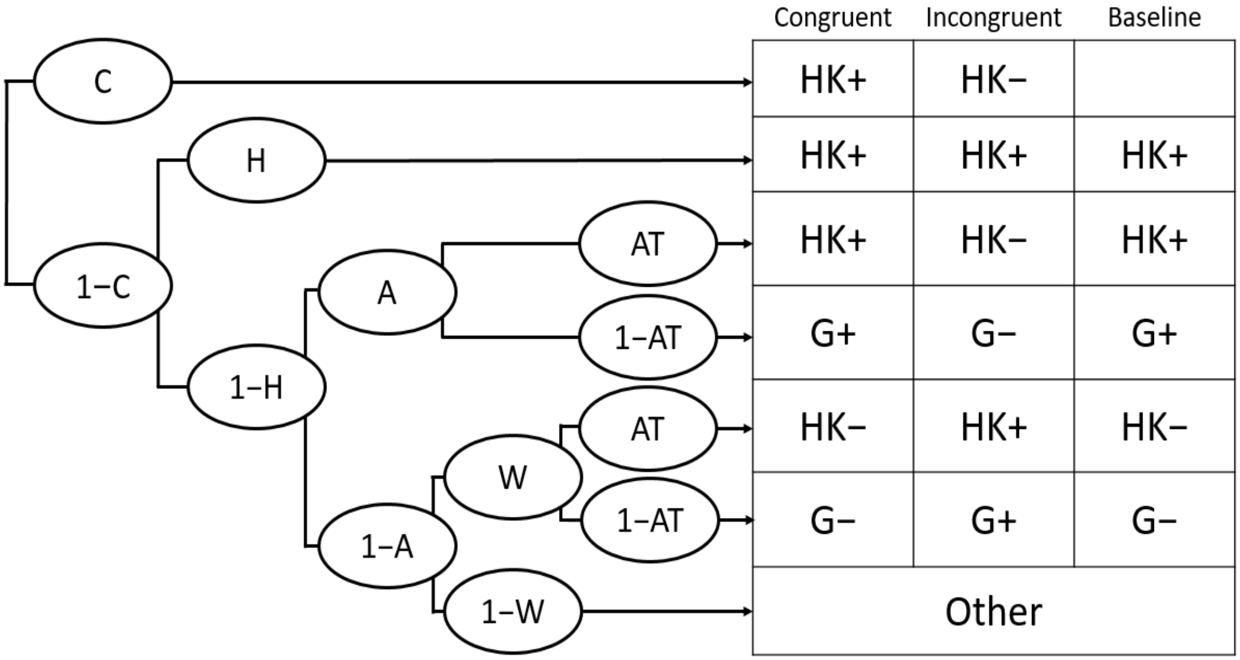

Figure 4. The capture model to be used in Experiment 1. The table indicates performance in the three context conditions (congruent/incongruent/baseline). Correct responses are signified by a plus sign (+), incorrect responses are signified by a minus sign (−), and “Other” refers to responses other than the target word or the pre-determined alternative word. Responses are characterized as either “heard” or “known” (HK) or as “guessed” (G).

I first ran two versions of this model to assess the role of inhibitory control (i.e., capture) in the SPIN and memory tasks. One model constrained the capture parameter to 0 and the other allowed the capture parameter to be estimated freely. The fit of these models to the data was then compared using the Bayesian information criterion (BIC) to determine whether adding the capture parameter to the model improved model fit. If adding the capture parameter improved model fit, this would indicate that having a means of achieving an accurate response in the congruent condition or a false hearing/memory response in the incongruent condition other than those provided by the other parameters in the model better explained the data. This would provide evidence for the role of inhibitory control – or another cognitive process that would result in the same response pattern – in these responses. For these models, all parameters were assumed to be equal across groups.

Next, multiple versions of the model including the capture parameter were fit to assess group differences in individual parameters. The simplest model assumed no group differences in any of the model's parameters. Increasingly complex iterations of the model allowing all combinations of parameters to vary between age groups were fit to the data, and model fits were compared using BIC values to determine which model best described the data.

In addition to BIC values, the goodness-of-fit of models to the data was assessed by comparing G^2 values to a critical G^2 value. As in Millar et al. (2018), I performed a compromise power analysis to set the critical value for each significance test using G*Power 3.1 (Faul et al., 2009). This compromise power analysis was done to compensate for the hypersensitivity of MPT models to detecting minor deviations between the observed values and those predicted by the model, which results in models being deemed to be statistically "poor fits" to the data when they, in fact, describe the observed data quite accurately (see Millar et al., 2018). Following the procedure used by Millar et al., the minimum effect size for the compromise power analysis was set at $\omega = .10$, the β/α ratio was set at 1.0, and the total sample size was calculated as the number of participants multiplied by the number of trials (SPIN task: 118 participants x 120 trials = 14,160; memory task: 118 participants x 90 trials = 10,620). When testing the fit of an individual model, the degrees of freedom was the number of independent response categories (i.e., all of the different possible response types in each context condition, such as baseline hits judged as guessed or cases of false memory in the incongruent condition judged as remembered or known) minus the number of parameters estimated. When comparing the fit of two models, the degrees of freedom was the number of models being compared minus one (2 – 1 = 1).

As in Jacoby et al. (2005), the accessibility bias and word generation parameters in all the models described above were allowed to differ in the baseline condition relative to the congruent and incongruent conditions. This was done under the assumption that accessibility of the target word and the likelihood of generating the pre-selected alternative word would differ in cases where context was unavailable relative to when it was available. However, unlike in the models created by Jacoby et al., I did not constrain the accessibility bias parameter to .50 in the baseline condition. Jacoby et al. set the accessibility bias parameter to .50 under the assumption that there should be a 50% chance of generating the target word and a 50% chance of generating the alternative word when the subject was unable to recollect the target word and there were no contextual cues available. However, this explanation failed to account for the fact that words other than the target word or the alternative word could be generated to complete the fragment. For example, for the cue-fragment pair head – s--l-, participants could generate the target word (skull), the alternative word (scalp), or a different word (e.g., smile). Thus, the probability of generating the target word when recollection fails should be less than .50 assuming an equal probability of generating all possible words in the absence of contextual cues. This issue is even more apparent in the hearing task, where any word could plausibly complete the sentence in the baseline condition (*The word is…*). For this reason, I allowed the accessibility bias parameter in the baseline condition to be estimated freely rather than setting it to .50, although as mentioned above, I allowed this parameter to differ in the baseline condition relative to the congruent and incongruent conditions. This decision was justified by the accessibility bias parameter in the baseline condition being estimated at .36 in the memory task and .16 in the SPIN task when all parameters were held constant across age groups.

2.2 Results

2.2.1 Stroop task

Before analyzing data from the SPIN and memory tasks to address my main hypotheses, I conducted a mixed-effects linear regression model using the glmer function from the *lme4* package in R (Bates et al., 2015) predicting reaction time within the auditory Stroop task to ensure that participants experienced the typical Stroop effect and to determine whether the magnitude of the Stroop effect differed across age groups, indicating the expected age difference in inhibitory control (Cohn et al., 1984; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West & Alain, 2000). This model was run on a subset of the full data including only baseline and incongruent trials, the trial types used by Sommers and Huff (2003, Experiment 2) to calculate the Stroop effect. Predictors in the model were an intercept term representing reaction time in the baseline condition for younger adults, a trial type variable representing the change in reaction time from the baseline condition to the incongruent condition (i.e., the magnitude of the Stroop effect) for younger adults, an age group variable representing the change in reaction time on baseline trials from younger to older adults, and the interaction of age group and trial type representing the change in the magnitude of the Stroop effect from younger to older adults. The reaction time in the baseline condition and the change in reaction time from the baseline condition to the incongruent condition were allowed to vary randomly across subjects, and reaction time was allowed to vary randomly across items (i.e., audio files for each speaker/word combination). For these analyses, I used only trials in which participants responded accurately (mean accuracy was .99 in the baseline condition and

.98 in the incongruent condition). Additionally, I excluded reaction times that were two standard deviations above or below the participants average in that trial type. On average, 5% of trials were excluded in each age group for this reason.

Average reaction times for accurate responses in the baseline and incongruent conditions of the auditory Stroop task are presented in Figure 5. The interaction of age group and trial type was non-significant, so it was removed from the model to allow for interpretation of main effects. This indicates that there was no difference in the magnitude of the Stroop effect across groups. The model revealed a robust Stroop effect in both age groups, as exhibited by a main effect of Stroop condition (*Estimated difference* [*ED*] = 73.24 ms, $t = 2.72$, $p < .05$). Additionally, older adults were slower to respond than younger adults in the baseline condition (*ED* = 201.16 ms, $t = 5.64$, $p < .001$). Thus, both groups experienced the predicted Stroop effect, and although older adults experienced general slowing as indicated by slower reaction times in the baseline condition, older adults did not experience an inflated Stroop effect as reported in previous studies (Cohn et al., 1984; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West & Alain, 2000)[3]. The lack of age differences in the magnitude of the Stroop effect foreshadows that inhibitory control as assessed by the auditory Stroop task may not be a good predictor of false hearing or false memory given that older adults are typically more susceptible than younger adults to false hearing (Rogers et al., 2012; Sommers et al., 2015) and false memory (Jacoby et al., 2005, Experiment 2).

---

[3] The same results were found when controlling for individual differences in hearing acuity by including best-ear PTA and the interaction of condition and best-ear PTA as predictors.

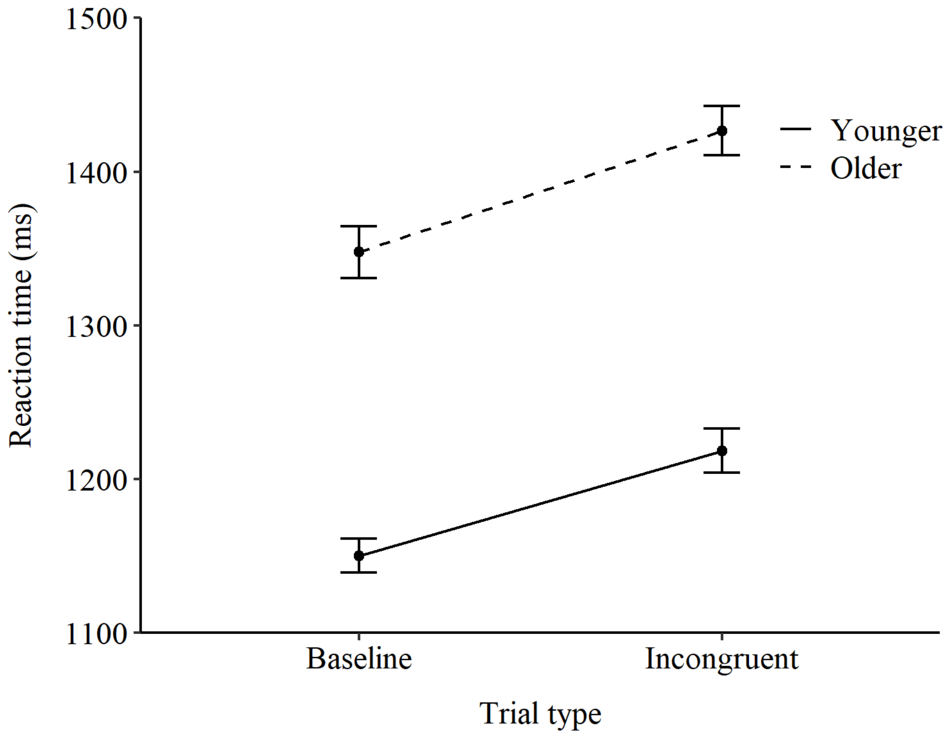

Figure 5. Average reaction times for accurate responses in the baseline and incongruent conditions of the auditory Stroop task by younger and older adults. Error bars represent 95% confidence intervals.

2.2.2 Executive functioning battery

Next, I conducted analyses to determine whether the five measures in the executive functioning battery tapped into a single construct. Descriptive statistics for each of the tasks in the executive functioning battery and results of *t*-tests comparing group means are presented in Table 1. Older adults achieved fewer categories in the Wisconsin Card Sorting Test, $t(61.95) = 5.39$, $p < .001$, and displayed poorer working memory capacity in the backward digit span than younger adults, $t(115.28) = 4.40$, $p < .001$, but did not differ from younger adults in the other tasks, *p*s > .05.

Table 1
Descriptive statistics for tasks comprising the executive functioning battery

	Younger adults		Older adults		
	M	*SD*	*M*	*SD*	*t-test*
WCST	5.97	.26	5.09	1.22	***
Arithmetic score	11.90	2.99	12.62	3.80	*n.s.*
Mental Control	23.07	4.12	22.79	3.95	*n.s.*
Backwards Digit Span	9.05	2.67	7.00	2.38	***
COWAT	47.62	10.76	46.31	13.90	*n.s.*

Note. WCST = Categories achieved in Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; *** = $p < .001$; *n.s.* = $p > .05$

I next sought to determine whether the individual components of the battery were correlated. As can be seen in Table 2, there were significant correlations between each component of the battery with the exception of the number of categories completed in the Wisconsin Card Sorting Test, which was only significantly related to the Arithmetic score from the WAIS-R and the Backwards Digit Span. The reason that the Wisconsin Card Sorting Test did not correlate with the other tasks in the battery was likely that there was limited variability in the number of categories that participants achieved. As can be seen in Table 1, nearly all younger adults completed all six categories, as did most older adults. Thus, the lack of significant correlations between the categories achieved on the Wisconsin Card Sorting Test and the other tasks in the battery likely resulted from limited individual differences.

Table 2
Bivariate correlations between components of executive functioning battery and auditory Stroop interference

	WCST	Arithmetic score	Mental Control score	Backwards Digit Span	COWAT	Stroop Diff
WCST	1.00					
Arithmetic score	.20*	1.00				
Mental Control score	.00	.28**	1.00			
Backwards Digit Span	.23*	.31***	.22*	1.00		
COWAT	.14	.19*	.32***	.34***	1.00	
Stroop Diff	-.14	.04	.03	-.05	.03	1.00

Note. WCST = Categories achieved in Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; Stroop Diff = Difference in reaction time for accurate responses between incongruent and baseline Stroop trials; * $p < .05$; ** $p < .01$; *** $p < .001$

To supplement the basic correlational analysis, I conducted an exploratory factor analysis (EFA) with varimax rotation using the factanal function from the *stats* package in R (R Core Team, 2012) to determine whether all five measures in the executive functioning battery loaded onto a single latent factor. Parallel analysis (Horn, 1965) was used to determine the optimal number of factors to retain using the parallel function in the *nFactors* package in R (Raiche & Magis, 2020)[4]. In parallel analysis, Monte Carlo simulations are conducted to create a distribution of eigenvalues from randomly generated datasets, then the eigenvalues from the EFA are compared to this distribution to determine whether the obtained eigenvalues were greater

---

[4] Parallel analysis has been shown to be a better method of factor extraction than Kaiser's (1960) method of retaining any factor with an eigenvalue greater than 1.00, which tends to overestimate the number of factors to retain (Zwick & Velicer, 1986).

than 95% of generated eigenvalues (i.e., significant at the $\alpha = .05$ level). If the EFA revealed that a single-factor model could explain the data and that all five measures within the executive functioning battery displayed high loadings on the single factor, then this would justify combining the measures to create a composite score.

Parallel analysis determined that the first factor was significant (eigenvalue = 1.92, critical = 1.15) and the second factor did not account for significant additional variance (eigenvalue = 1.03, critical = 1.04). As can be seen in Table 3, each of the measures in the executive functioning battery demonstrated moderate-to-strong loadings on the single factor, although the Wisconsin Card Sorting Test displayed a weaker loading likely due to the limited variability mentioned above. Based on the results of the correlations and the EFA, I felt justified in combining the tasks into a single executive functioning composite score as in past studies (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007).

Table 3
Factor loadings from the EFA of measures in the executive functioning battery

	Factor loading
WCST	.29
Arithmetic score	.49
Mental Control score	.45
Backwards Digit Span	.62
COWAT	.54

Note. WCST = Categories achieved on the Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test

Having created the executive functioning composite, I next tested whether younger and older adults differed in executive functioning. As expected, older adults (*M* = -.16, *SD* = .68) exhibited poorer executive functioning than younger adults (*M* = .16, *SD* = .51), *t*(105.8) = 2.93, *p* < .01. This converges with past research showing that executive functioning declines with age (Chan & McDermott, 2007; McCabe et al., 2009; Roediger & McDaniel, 2007).

Finally, I conducted analyses to determine whether inhibitory control as assessed by the auditory Stroop task was related to the measures in the executive functioning battery. As can be seen in Table 2, the difference in reaction time for accurate responses between incongruent and baseline Stroop trials did not correlate significantly with any of the measures of executive functioning. This suggests that the aspect of inhibitory control captured by interference on the auditory Stroop task was not related to performance on any individual component of the executive functioning battery.

To supplement the correlational analysis, I conducted a second EFA to determine whether inhibitory control as assessed by the auditory Stroop task loaded onto the same factor as the measures in the executive functioning battery. Since inhibitory control is typically considered an executive function (Jurado & Rosselli, 2007), inhibitory control as assessed by the auditory Stroop task should load onto the same factor as the measures of executive functioning. It was also important to establish that inhibitory control as assessed by the auditory Stroop task loaded onto the executive functioning factor for the planned analysis described below in which the executive functioning score is added to a mixed-effects logistic regression model including age group, baseline Stroop reaction time, and incongruent Stroop reaction time to determine whether executive functioning was predictive of false hearing and false memory after controlling for the response suppression aspect of inhibitory control. If inhibitory control as assessed by the auditory Stroop task did not load onto the same factor as measures of executive functioning, then this would suggest that the auditory Stroop task measured something other than what was included in the executive functioning battery – either a different component of inhibitory control (Kramer et al., 1994; Lustig et al., 2007) or something other than inhibitory control – and I would not be able to draw the conclusion that the inhibitory control component of the executive functioning battery was controlled for in this analysis. The EFA included each of the measures from the executive functioning battery and a standardized difference score subtracting baseline Stroop reaction time from incongruent Stroop reaction time. The difference score was calculated to demonstrate differences in inhibitory control accounting for variability in processing speed.

Parallel analysis revealed a significant first factor (eigenvalue = 1.93, critical = 1.20) and a significant second factor (eigenvalue = 1.16, critical = 1.08). As can be seen in Table 4, the Stroop difference score only loaded onto the second factor, and the Wisconsin Card Sorting Test loaded strongly onto this factor. However, the Stroop difference score only loaded weakly onto this factor, suggesting that it shared little variance with either latent factor. All the other measures in the executive functioning battery loaded strongly onto the first factor and only weakly onto the second factor. The findings of the EFA in addition to the low correlations between the auditory Stroop difference score and each measure in the executive functioning battery suggest that inhibitory control as assessed by the auditory Stroop task tapped a latent variable that was different from that tapped by measures in the executive functioning battery[5]. Therefore, the analysis proposed above including age group, baseline Stroop reaction time,

---

[5] I conducted this analysis again using incongruent Stroop reaction time as an index of inhibitory control instead of the Stroop difference score and obtained similar results. The data indicated a second significant factor (eigenvalue = 1.24, critical = 1.08), and both WCST and incongruent Stroop reaction time loaded onto one factor whereas the other measures of executive functioning loaded onto the other factor.

incongruent Stroop reaction time, and executive functioning score as simultaneous predictors can be used to determine whether the executive functioning score made an independent contribution to predicting false memory and false hearing controlling for the other variables, but it cannot be concluded that controlling for Stroop task performance fully partialled out the inhibitory control component of the executive functioning battery. This issue will be discussed further below.

Table 4
Factor loadings from the EFA of measures in the executive functioning battery and the Stroop difference score

	Factor 1 loading	Factor 2 loading
WCST	.22	.57
Arithmetic score	.48	.12
Mental Control score	.57	-.22
Backwards Digit Span	.55	.23
COWAT	.55	
Stroop difference score		-.23

Note. WCST = Categories achieved in Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; Stroop difference score = difference in reaction time between the baseline and incongruent conditions of the auditory Stroop task

2.2.3 Memory task

2.2.3.1 Accuracy

Accuracy in the memory task was analyzed using mixed-effects logistic regression using the glmer function from the *lme4* package in R (Bates et al., 2015). The dependent variable in this model was trial-by-trial accuracy. The model included an intercept term corresponding to the odds of an accurate response for younger adults in the baseline condition, two dummy coded variables indicating the change in the odds of an accurate response from the baseline condition to the congruent and incongruent conditions for younger adults, a group variable representing the change in odds of an accurate response from younger to older adults in the baseline condition, and the interaction of group with the congruent and incongruent condition dummy codes to determine whether the change in the odds of an accurate response from the baseline condition to the congruent and incongruent conditions differed between younger and older adults. The odds of an accurate response in the baseline condition and the change in odds of an accurate response from the baseline to the congruent and incongruent conditions were allowed to vary randomly across subjects, and the odds of an accurate response were allowed to vary randomly across items (i.e., target words).

As shown in Figure 6, patterns in the memory task closely resembled those from Jacoby et al. (2005, Experiment 2). First, it is important to note that the age groups differed in baseline accuracy, with the odds of an accurate response being 1.69 times greater in the older adult group than in the younger adult group ($z = 4.15$, $p < .001$)[6]. As in Jacoby et al., however, I found that younger adults' performance improved in the congruent condition relative to the baseline condition ($OR = 5.96$, $z = 6.84$, $p < .001$), and that older adults experienced marginally greater improvement to performance from the baseline to the congruent condition relative to younger adults ($OR = 1.56$, $z = 1.96$, $p = .05$). Similarly, younger adults' performance was negatively impacted in the incongruent condition relative to the baseline condition ($OR = .47$, $z = -3.06$, $p < .01$), and older adults' performance was negatively impacted to a significantly greater degree than was that of younger adults ($OR =.33$, $z = -5.60$, $p < .001$). This supports the conclusion that

---

[6] Age groups demonstrated equivalent performance in the baseline condition of the same task in the study by Jacoby et al. (2005). It is possible that adding a 186 second retention interval for the younger adult group to compensate for the shorter presentation duration of stimuli in the study phase resulted in lower accuracy for younger, relative to older, adults in the present study.

older adults were more likely to respond based on available context (i.e., the prime word) than were younger adults.

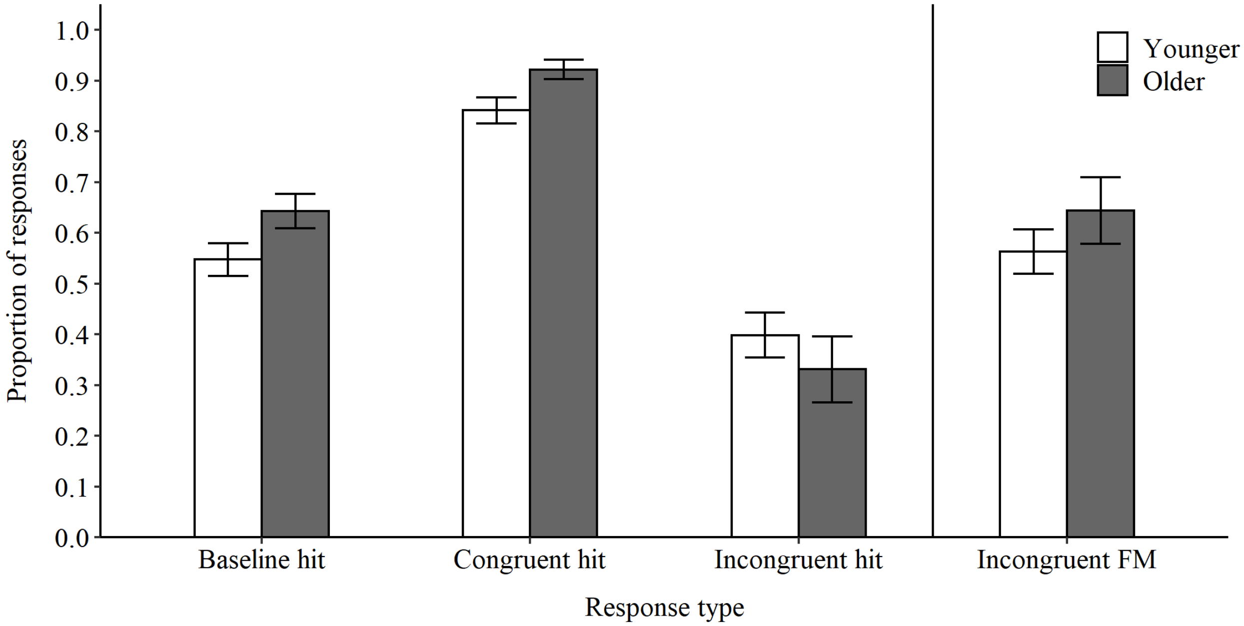

Figure 6. Average proportion of hits and susceptibility to false memory (FM) for younger and older adults in the memory task. Error bars represent 95% confidence intervals.

2.2.3.2 False memory

I then compared the odds of experiencing false memory across age groups in the incongruent condition of the memory task. For this analysis, I created a mixed-effects logistic regression model predicting trial-by-trial false memory on a subset of data that included only incongruent trials. The predictors in this model were an intercept term corresponding to the odds of a false memory response in the younger adult group and an age group variable corresponding to the change in the odds of a false memory response from the younger adult group to the older adult group. The odds of experiencing false memory was allowed to vary randomly across subjects and across items.

As in Jacoby et al. (2005, Experiment 2), older adults in Experiment 1 were more susceptible to false memory than younger adults (see Figure 6). For younger adults, the odds of experiencing a false memory was 1.39 times greater than the odds of giving any other response combined on incongruent trials, which was not a significant difference ($z = 1.62$, $p > .05$). The odds that an older adult would experience a false memory were 1.79 times greater than for younger adults ($z = 2.48$, $p < .05$). This further supports the idea that older adults were more likely to respond based on context, leading to false memory when context was misleading.

2.2.3.3 Subjective judgements of false memory

I next investigated the relative odds of rating false memories as remembered versus known or guessed across groups. A mixed-effects multinomial logistic regression analysis was conducted to assess group differences in susceptibility to false memory responses judged as remembered, known, and guessed. This analysis was conducted using the brm function from the *brms* package in R (Bürkner, 2017). The model separately generated the odds of giving a know or a guess classification relative to a remember classification. An intercept term corresponding to the odds of giving either a know or guess classification relative to a remember classification in younger adults and a dummy-coded grouping variable corresponding to the change in those odds from younger to older adults were included as predictors. The relative odds of experiencing a false memory rated as remembered versus known or guessed was allowed to vary randomly across subjects and items. The brm function does not output *p*-values, but the statistical significance of estimates can be assessed by the 95% confidence interval (*95% CI*) surrounding the estimate: If the confidence interval includes 1.00, then the estimated odds ratio is not statistically significant at the $\alpha = .05$ level.

As can be seen in Figure 7, younger adults were numerically least likely to rate false memories as remembered and most likely to rate false memories as guessed. However, there was no significant difference in the odds of characterizing a false memory as known (*OR* = 1.18, *95% CI* = .73 – 1.95) or guessed (*OR* = 1.35, *95% CI* = .78 – 2.37) relative to remembered for younger adults. Older adults displayed the opposite pattern, rating most cases of false memory as remembered and the fewest as guessed. The difference in pattern between younger and older adults was reflected in a significant group difference in the change in odds from a remember rating to a guess rating (*OR* = .16, *95% CI* = .08 – .33). The change in odds from a remember rating to a know rating did not differ significantly across age groups (*OR* = .52, *95% CI* = .26 – 1.04). These findings suggest that younger adults had better metacognition regarding the inaccuracy of their false memory experiences than did older adults.

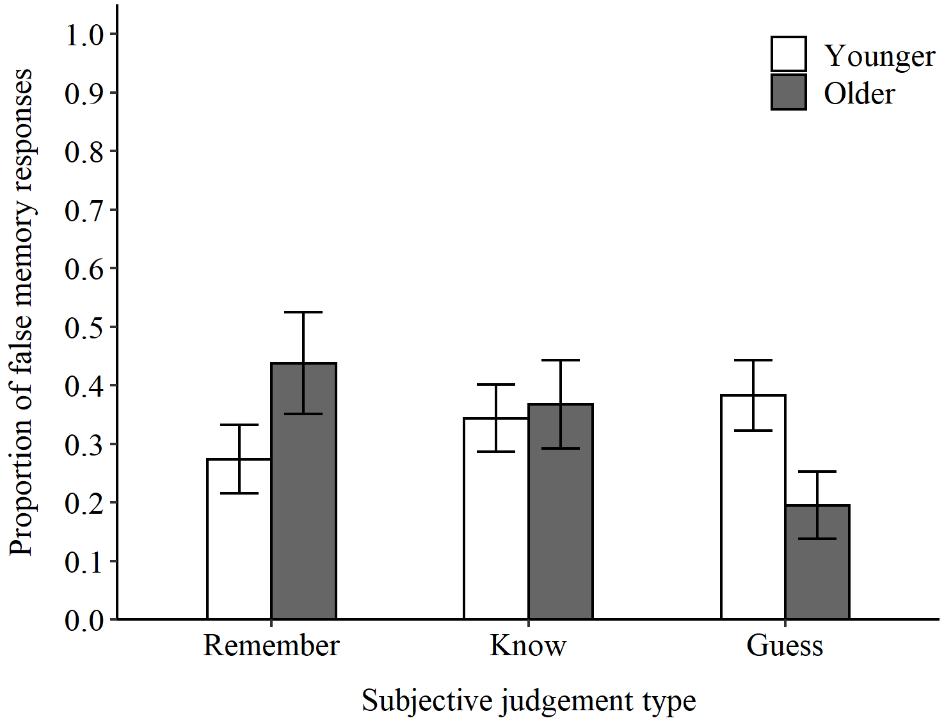

Figure 7. Proportion of false memory responses characterized as remembered, known, or guessed for younger and older adults. Error bars represent 95% confidence intervals.

2.2.3.4 Relationships with inhibitory control and executive functioning

To test the effects of the response suppression aspect of inhibitory control and executive functioning, I conducted separate mixed-effects logistic regression models predicting trial-by-trial false memory. In the inhibitory control models, average reaction times from the baseline condition of the auditory Stroop task standardized across all participants and age group were entered as simultaneous predictors. I then added standardized average incongruent Stroop reaction times to the model and compared it to the previous model using the anova function in R to determine whether incongruent Stroop reaction time accounted for additional variance after controlling for age group and general processing speed (i.e., baseline Stroop reaction time). The executive functioning model included the executive functioning composite score standardized across all participants and age group as predictors to determine whether executive functioning was predictive of false hearing or false memory after controlling for age group. I also tested the interaction between age group and the cognitive predictors (baseline and incongruent Stroop reaction times and the executive functioning score). However, none of the interaction effects were significant, so they were removed from the models to allow for interpretation of the main effects. The odds of experiencing false memory were allowed to vary randomly across subjects and items in these models.

To test the relationship between the response suppression aspect of inhibitory control and false memory, I first conducted the base model including age group and average baseline Stroop reaction time standardized across groups as predictors. In this model, age group was marginally significant, with the odds of experiencing a false memory being 1.62 times greater in older adults than in younger adults controlling for average processing speed ($z = 1.82$, $p = .07$). Processing speed was not a significant predictor of false memory after controlling for age group ($OR = 1.15$, $z = .84$, $p > .05$). Adding incongruent Stroop reaction time did not improve the model ($\chi^2 = .67$, $p > .05$), indicating that the response suppression aspect of inhibitory control was not a significant predictor of false memory after controlling for age and processing speed. Age group remained a marginally significant predictor of susceptibility to false memory controlling for average baseline and incongruent Stroop reaction time ($OR = 1.58$, $z = 1.72$, $p = .08$), but the effects of baseline and incongruent Stroop reaction times were not significant (*p*s > .05).

The relationship between executive functioning and false memory, however, was significant. After controlling for age group, a one standard deviation increase in executive functioning score decreased the odds of experiencing a false memory by a factor of 2.29 ($z = -4.46$, $p < .001$). Age group did not significantly predict false memory after controlling for executive functioning (*OR* = 1.38, *z* = 1.44, *p* > .05), suggesting that differences in executive functioning could fully account for age differences in false memory. Additionally, the full model including group, baseline and incongruent Stroop reaction times, and executive functioning score was a significantly better fit to the data than one that included only group and baseline/incongruent Stroop reaction times ($\chi^2 = 18.34$, $p < .001$). Whereas executive functioning score was a significant predictor of false memory in this full model (*OR* = .44, *z* = -4.45, *p* < .001), none of the other predictors were significant (*p*s > .05). This indicates that executive functioning accounted for significant variance in susceptibility to false memory after controlling for age, processing speed, and the response suppression aspect of inhibitory control as assessed by the auditory Stroop task. However, as discussed above, the results of an EFA suggested that inhibitory control as assessed by the auditory Stroop task tapped into a different latent construct than the executive functioning battery. Therefore, it cannot be concluded from this analysis that executive functions other than the response suppression aspect of inhibitory control were significantly related to susceptibility to false memory since the inhibitory control component of the executive functioning battery may not have been fully partialled out by including incongruent Stroop performance in the model.

2.2.4 SPIN task

Statistical analyses for the SPIN task were nearly identical to those for the memory task. The only difference was that best-ear PTA was included as a covariate in analyses testing the relationships between the response suppression aspect of inhibitory control, executive functioning, and false hearing. This was done to further control for individual differences in hearing acuity above and beyond the between-groups SNR manipulation, as individual differences in hearing acuity may account for differences in susceptibility to false hearing.

2.2.4.1 Accuracy

The patterns in the SPIN task closely resembled those in past studies of false hearing (see Figure 8; Rogers et al., 2012; Sommers et al., 2015). The SNR manipulation – i.e., using a -4 dB SPL SNR for younger adults and a more favorable +1 dB SPL SNR for older adults to account for age differences in hearing acuity (Morrell et al., 1996; Sommers et al., 2011) – did not successfully equate performance in the baseline condition across age groups. The odds of an accurate response in the baseline condition were 1.52 times greater in older than younger adults ($z = 2.22$, $p < .05$), suggesting that the more favorable SNR allowed older adults to more accurately identify target words in the absence of predictive context. Younger adults were significantly more likely to accurately identify the target word in the congruent condition relative to the baseline condition ($OR = 15.96$, $z = 7.55$, $p < .001$), demonstrating benefit to speech perception from the presence of valid contextual cues. Despite being better able to identify the target words in the baseline condition, older adults experienced significantly greater improvement from the baseline to the congruent condition relative to younger adults ($OR = 1.71$, $z = 2.34$, $p < .05$). In the incongruent condition, younger adults were less likely to accurately identify the target word than in the baseline condition ($OR = .26$, $z = -3.75$, $p < .001$), and older adults experienced significantly greater detriment to performance from the baseline to the incongruent condition than did younger adults ($OR = .62$, $z = -2.17$, $p < .05$). Thus, as in the memory task, older adults' performance was influenced to a greater degree by the presence of valid and invalid context relative to younger adults.

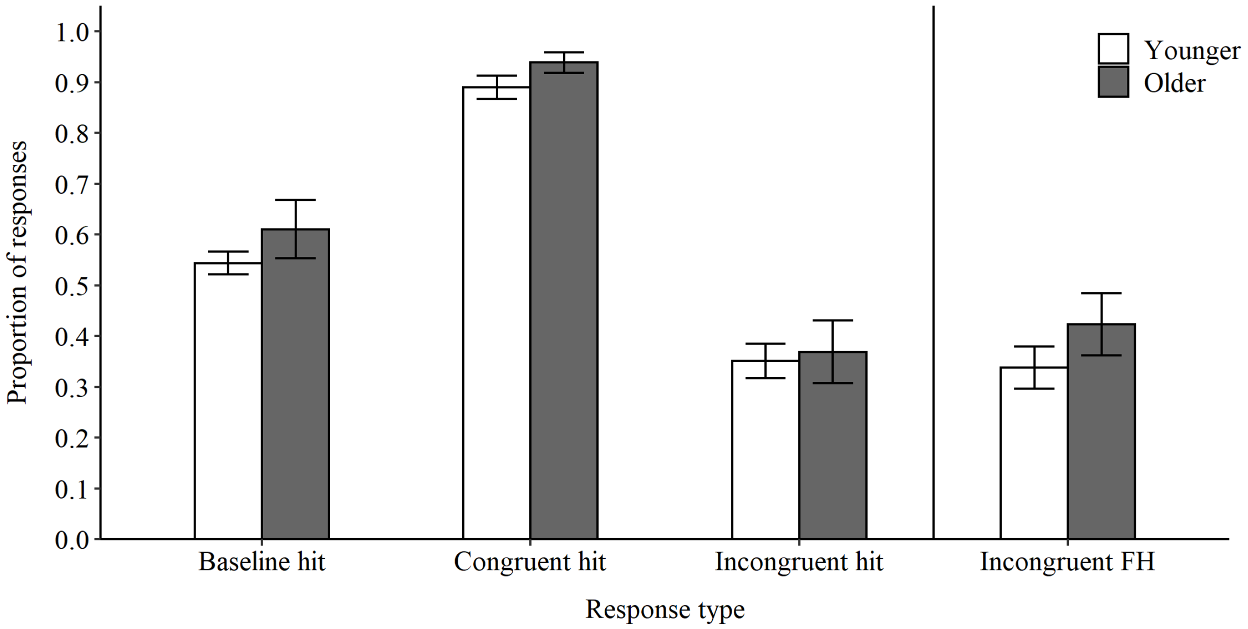

Figure 8. Average proportion of hits and susceptibility to false hearing (FH) for younger and older adults in the SPIN task. Error bars represent 95% confidence intervals.

2.2.4.2 False hearing

As can be seen in Figure 8, younger and older adults both experienced false hearing in the incongruent condition. Unlike for false memory, the odds of experiencing false hearing was 3.16 times lower than the odds of giving any other response combined on incongruent trials for younger adults (z = -3.86, p < .001). As in past studies (Rogers et al., 2012; Sommers et al., 2015), however, older adults were marginally more likely to experience false hearing than younger adults, with the odds of experiencing false hearing being 1.71 times greater in older, relative to younger, adults (z = 1.92, p = .05). Thus, as was true in the memory task, older adults were more likely than younger adults to erroneously use invalid contextual cues in speech perception, resulting in greater susceptibility to false hearing.

2.2.4.3 Subjective judgements of false hearing

Unlike in the memory task, both age groups displayed poor metacognition in the SPIN task, frequently characterizing cases of false hearing as heard. As can be seen clearly in Figure 9, younger adults were less likely to characterize cases of false hearing as guessed relative to heard (*OR* = .04, *95% CI* = .02 – .08). Younger adults were also numerically less likely to characterize cases of false hearing as known relative to heard, but this difference did not reach significance (*OR* = .64, *95% CI* = .39 – 1.04). The change in odds between hear and know judgements was significantly greater in older adults (*OR* = .43, *95% CI* = .25 – .74), although there was no group difference in the change of odds between hear and guess judgements (*OR* = .88, *95% CI* = .41 – 2.02). These findings suggest that both younger and older adults often had the subjective experience of hearing the contextually predicted word in incongruent sentences even though a different (but similar sounding) word was presented.

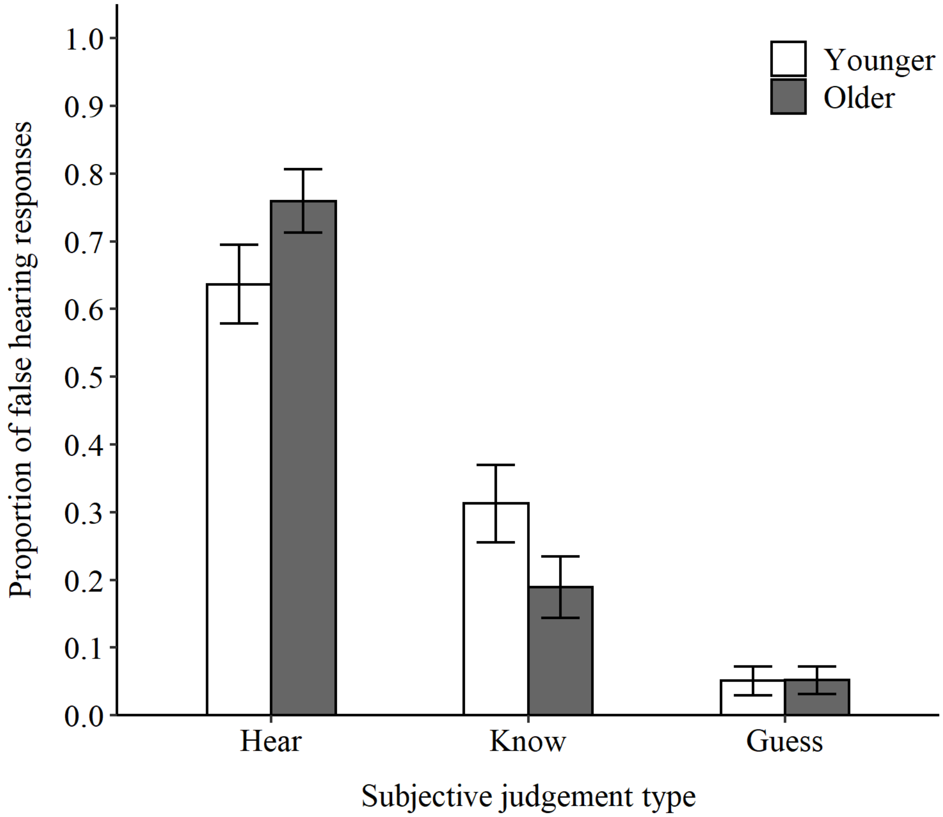

Figure 9. Proportion of false hearing responses characterized as heard, known, or guessed for younger and older adults. Error bars represent 95% confidence intervals.

2.2.4.4 Relationship with false memory

Before testing the relationship between false hearing and both inhibitory control and executive functioning, I first tested the direct relationship between false hearing and false memory. After all, if there was no relationship between false hearing and false memory, it may be a moot point to look for common mechanisms. Susceptibility to false hearing was positively correlated with susceptibility to false memory ($r = .39$, $p < .001$; see Figure 10). This indicates that individuals who were more susceptible to false memories also tended to be more susceptible to false hearing.

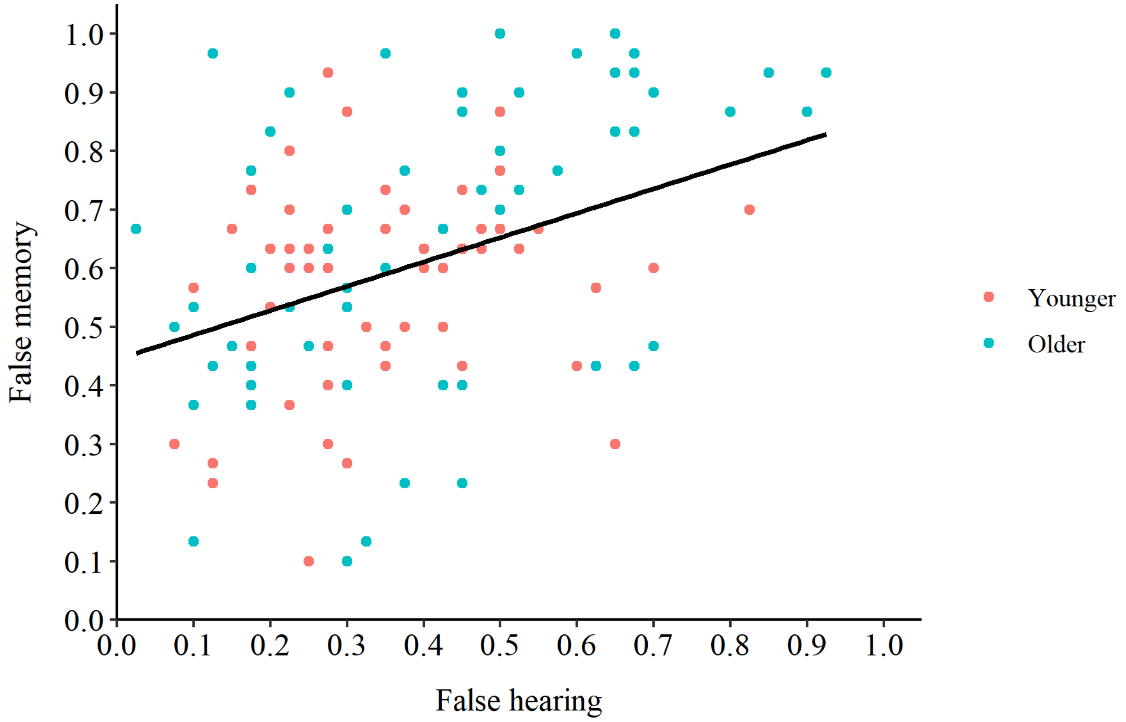

Figure 10. Correlation between average susceptibility to false hearing and average susceptibility to false memory ($r = .39$).

To supplement the basic correlational analysis, I conducted a mixed-effects logistic regression with age group, average susceptibility to false memory standardized across participants, and the interaction of age group and average susceptibility to false memory as predictors of trial-by-trial false hearing. The interaction of age group and average susceptibility to false memory was not significant, so it was removed from the model. The odds of experiencing false hearing were allowed to vary randomly across subjects and items.

As was true of the correlational analysis, the mixed-effects logistic regression model suggested that individuals who were more susceptible to false memory were also more susceptible to false hearing: A one standard deviation increase in average susceptibility to false memory increased the odds of false hearing by a factor of 1.97 after controlling for age group ($z = 5.19$, $p < .001$). Age group was not a significant predictor of false hearing after controlling for average susceptibility to false memory (*OR* = 1.33, z = 1.10, p > .05). The presence of a direct relationship between false hearing and false memory suggests that at least one common mechanism may underlie these phenomena. I next examined the relationships between false hearing and both the response suppression aspect of inhibitory control and executive functioning to determine whether similar relationships to those observed in the false memory analyses emerged.

2.2.4.5 Relationship with inhibitory control and executive functioning

In the base model assessing the response suppression aspect of inhibitory control, age group did not significantly predict susceptibility to false hearing after controlling for individual differences in processing speed and best-ear PTA (*OR* = .90, z = -.26, p > .05). However, the effect of average baseline Stroop reaction time approached significance (*OR* = 1.89, z = 1.66, p = .10), indicating that slower processing speed on the auditory Stroop task tended to be related to increased susceptibility to false hearing in younger adults controlling for best-ear PTA[7]. Additionally, as mentioned above, the interaction of age group and baseline Stroop reaction time approached significance (*OR* = .46, z = -1.75, p = .08), suggesting that the relationship between processing speed on the Stroop task and false hearing was weaker in older, relative to younger, adults. Best-ear PTA was the only significant predictor of false hearing in this model, with a one standard deviation increase in best-ear PTA increasing susceptibility to false hearing by a factor of 1.44 controlling for age group and processing speed (z = 2.01, p < .05). Of interest was whether adding average incongruent Stroop reaction time would account for additional variance

---

[7] This effect only pertains to younger adults because the model also included age group. The model controls for age group by assessing the effect of baseline Stroop reaction time when group equals 0. Since the younger adult group was coded 0, the effect of baseline Stroop reaction time pertains to the younger adult group, whereas the interaction of age group and baseline Stroop reaction time indicates how this relationship changed from the younger adult group to the older adult group.

in the data, which was not the case ($\chi^2 = 0.41$, $p > .05$). This indicates that the response suppression aspect of inhibitory control as assessed by the auditory Stroop task was not predictive of false hearing after controlling for age, processing speed, and best-ear PTA. All predictors were non-significant in this final model (*p*s > .05) aside from best-ear PTA, which was marginally significant (*OR* = 1.42, $z = 1.93$, $p = .05$).

As was true in false memory, executive functioning was a significant predictor of false hearing. After controlling for age group and best-ear PTA, a one standard deviation increase in executive functioning reduced the odds of false hearing by a factor of 2.53 ($z = -4.20$, $p < .001$). Age group was not a significant predictor of false hearing after controlling for executive functioning and best-ear PTA (*OR* = .90, $z = -.32$, $p > .05$), again suggesting that differences in executive functioning fully accounted for age-related differences in false hearing. Best-ear PTA was also non-significant after controlling for age group and executive functioning (*OR* = 1.31, $z = 1.54$, $p > .05$). Additionally, adding executive functioning to the model of inhibitory control including age group, baseline/incongruent Stroop reaction times, the group by baseline Stroop interaction, and best-ear PTA improved model fit ($\chi^2 = 18.94$, $p < .001$), indicating that executive functioning was predictive of susceptibility to false hearing even after controlling for age, processing speed, the response suppression aspect of inhibitory control as assessed by the auditory Stroop task, and hearing acuity. However, as discussed above, the results of an EFA suggested that inhibitory control as assessed by the auditory Stroop task tapped into a different latent construct than the executive functioning battery. Therefore, as with false memory, I cannot conclude from this analysis that executive functions other than the response suppression aspect of inhibitory control were significantly related to susceptibility to false hearing. In this model, both executive functioning (*OR* = .38, $z = -4.51$, $p < .001$) and the group by baseline Stroop reaction time interaction (*OR* = .41, *z* = -2.02, *p* < .05) were significant, but all other predictors were non-significant (*p*s > .05).

To determine whether the relationship between false hearing and false memory could be fully accounted for by individual differences in executive functioning, I tested whether the relationship between false hearing and false memory remained significant after controlling for executive functioning. For this analysis, I conducted a new mixed-effects logistic regression model with age group, average susceptibility to false memory, and score on the executive functioning battery as predictors of false hearing. Susceptibility to false hearing was allowed to vary randomly across subjects and items. If the relationship between false hearing and false memory remained significant after controlling for executive functioning, this would indicate that factors above and beyond those captured by the executive functioning battery contributed to the relationship between false hearing and false memory. In this model, both executive functioning (*OR* = .52, *z* = -2.88, *p* < .01) and false memory (*OR* = 1.70, *z* = 3.87, *p* < .001) were independently predictive of false hearing after controlling for age group. Age group was not significant after controlling for average susceptibility to false memory and executive functioning score (*OR* = 1.14, *z* = .53, *p* > .05). This suggests that despite being predictive of both false hearing and false memory, executive functioning cannot fully account for the relationship between these two phenomena. Additionally, it suggests that the executive functioning battery accounted for unique variance in false hearing that was not accounted for by susceptibility to false memory. Therefore, it is possible that aspects of the executive functioning battery tapped into variability in susceptibility to false hearing related to individual differences in speech perception or sentence processing abilities that were not captured by average susceptibility to false memory.

2.2.5 MPT modeling

2.2.5.1 Memory task

Beginning with the memory task, I found that neither the model in which the capture parameter was constrained to 0 ($G^2(18) = 475.17$, critical = 52.75, BIC = 530.79) nor the model in which the capture parameter was freely estimated ($G^2(17) = 425.08$, critical = 51.49, BIC = 489.97) – which I will refer to simply as the capture model – provided good fits to the data. However, the capture model did provide significantly better fit to the data than did the model that did not include the capture parameter ($\Delta G^2(1) = 50.09$, critical = 27.23). This, in addition to the lower BIC value in the capture model than in the model without capture, suggests that capture – interpreted as representing inhibitory control – influenced responding in the memory task. This finding may seem surprising given that inhibitory control as assessed by the auditory Stroop task was unrelated to susceptibility to false memory. Potential reasons for this discrepancy are outlined in the Experiment 1 discussion.

Since MPT modeling can be overly sensitive to small deviations between the observed and estimated values when determining model fit (see Millar et al., 2018), I plotted the observed and predicted values for the capture model to determine whether its fit was as poor as the G^2 value would suggest (see Figure 11). The plot suggests that the capture model generally did a good job of predicting responses in the memory task, although there were some deviations for hits judged as remembered/known in the baseline and congruent conditions and false memory responses in the incongruent condition. However, it is important to note that this was the most basic model in which none of the parameters were allowed to deviate between younger and older adults. Since younger and older adults differed in their responding in the memory task, we would expect better fit for models that account for age differences.

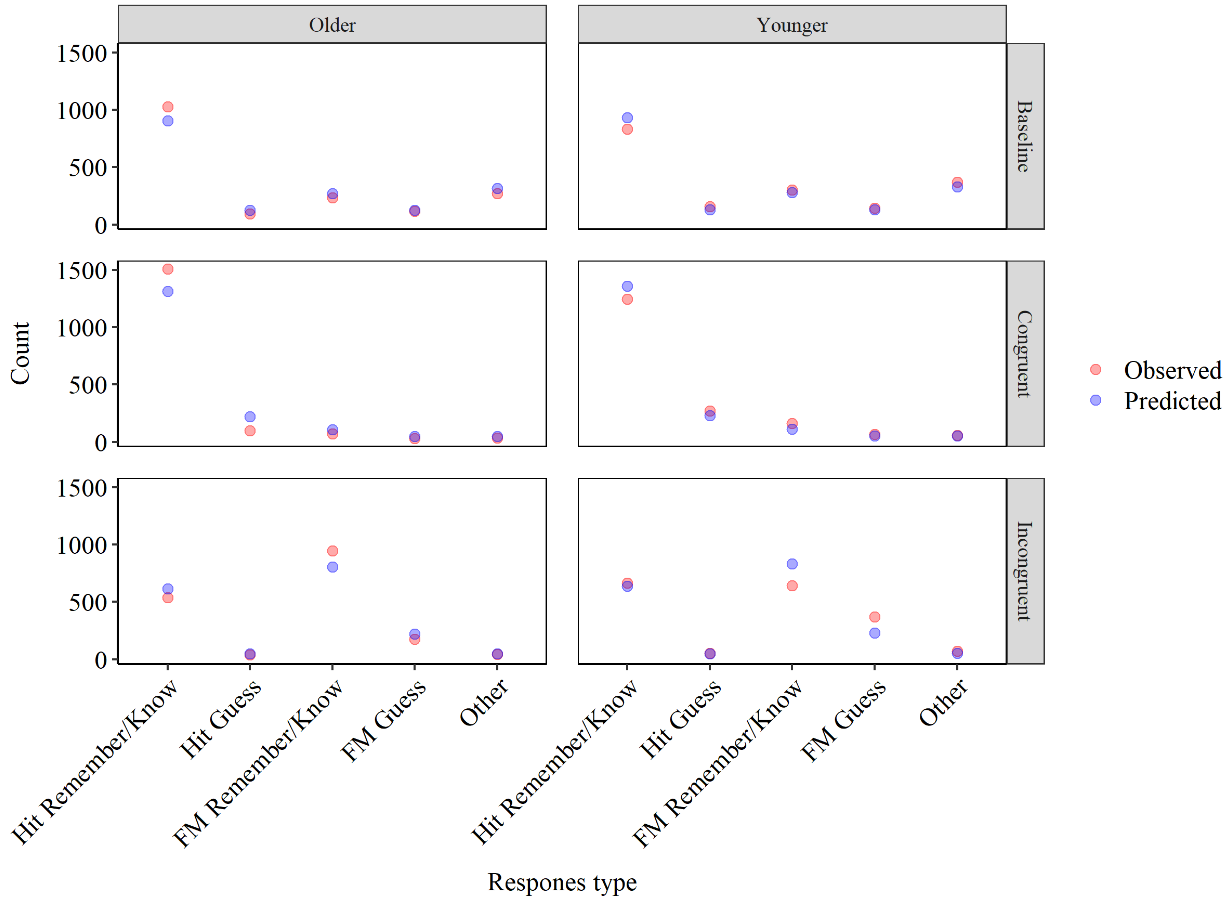

Figure 11. Observed versus predicted counts of responses in each context condition for the base capture model of memory. FM = false memory; Other = responses other than the target word or the pre-selected alternative word.

To assess whether individual parameters in the capture model differed between age groups, I created models in which each individual parameter was allowed to vary between age groups while the other parameters were held constant and compared these models to the base capture model in which all parameters were held constant. I found significant age differences in capture ($\Delta G^2(1) = 245.59$, critical = 27.22, BIC = 253.66), recollection ($\Delta G^2(1) = 78.43$, BIC = 420.81), accessibility bias ($\Delta G^2(1) = 60.61$, BIC = 447.90), and attribution threshold ($\Delta G^2(1) = 176.58$, BIC = 322.66), but not in word generation ($\Delta G^2(1) = .61$, BIC = 507.90). Looking at the individual parameter estimates revealed that older adults were more likely to be captured by context than younger adults (.37 vs. .01), were more likely to recollect the target word than younger adults (.46 vs. .32), had greater accessibility bias than younger adults (in baseline condition: .44 vs. .34; in congruent/incongruent conditions: .85 vs. .74), and were more likely to characterize their responses as remember or know than younger adults (.80 vs. .62). While the finding that older adults were deemed more likely to recollect the target word than younger adults may seem to contradict past findings of age-related declines in memory (for a review, see Nyberg et al., 2012), this merely indexes older adults' better performance in the baseline condition of the memory task relative to younger adults. Since the presentation-time manipulation in the study phase (i.e., showing older adults the word pairs for 3000 ms vs. 1000 ms for younger adults) resulted in better performance by older, relative to younger, adults in the present study, this difference was reflected in the MPT model.

Finally, I created models allowing all possible combinations of parameters to differ across younger and older adults to determine which model best fit the memory task data. Table 5 lists the G^2 and BIC values for each of these models. Comparing these models based on their BIC values, the best fitting model allowed for group differences in capture and recollection. This model is parsimonious, suggesting that only two of the five parameters needed to be allowed to differ across age groups to best account for the observed data. It should be noted that this model was still classified as a poor fit to the data according to its G^2 value ($G^2(15) = 63.91$, critical = 48.93). However, as can be seen in Figure 12, the model did a much better job of predicting responses than the most basic capture model, with only minor deviations between the observed and predicted response counts. Thus, I believe that the model accounting for age differences in capture and recollection is useful for describing the data from the memory task.

Table 5

Fit statistics for each MPT model in the memory task

Parameters allowed to vary	G^2	BIC
None	425.08	489.97
C	179.49	253.66
R	346.65	420.81
A	364.47	447.9
W	424.47	507.9
AT	248.5	322.66
C, R	63.91	147.34
C, A	164.24	256.94
C, W	178.89	271.59
C, AT	138.48	221.91
R, A	312.68	405.38
R, W	344.61	437.32
R, AT	239	322.44
A, W	362.7	464.67
A, AT	160.86	253.56
W, AT	247.86	340.57
C, R, A	57.73	159.7
C, R, W	61.38	163.36
C, R, AT	62.97	155.68
C, A, W	163.67	274.92
C, A, AT	113.45	215.42
C, W, AT	137.91	239.88
R, A, W	308.48	419.72
R, A, AT	160.79	262.77
R, W, AT	238.4	340.37
A, W, AT	159.77	271.02
C, R, A, W	55.74	176.25

C, R, A, AT	57.47	168.71
C, R, W, AT	60.7	171.95
R, A, W, AT	159.75	280.27
C, R, A, W, AT	55.35	185.13

Note. BIC = Bayesian Information Criterion; C = Capture; R = Recollection; A = Accessibility bias; W = Word generation; AT = Attribution threshold.

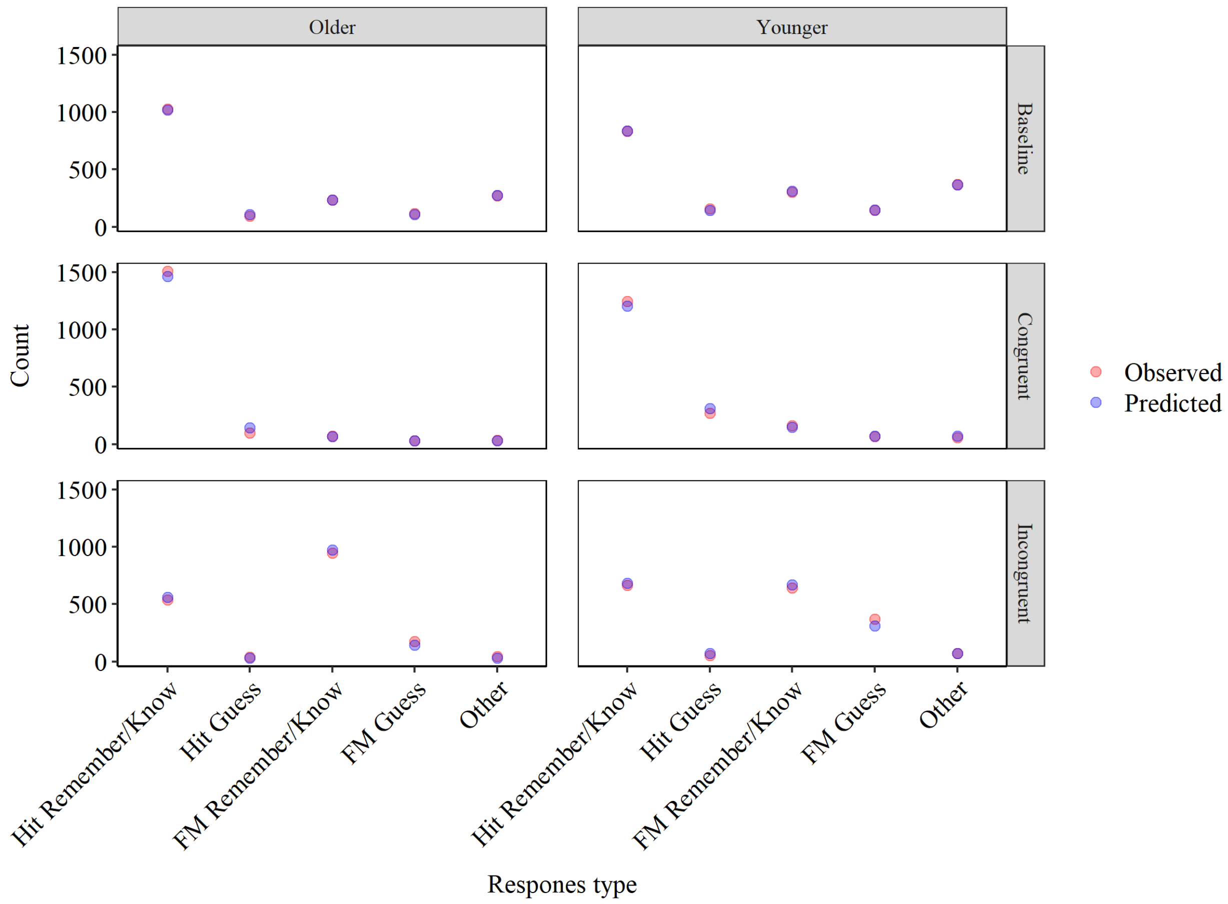

Figure 12. Observed versus predicted counts of responses in each context condition for the capture model of memory allowing group differences in capture and recollection. FM = false memory; Other = responses other than the target word or the pre-selected alternative word.

2.2.5.2 SPIN task

Turning next to the SPIN task, I found that neither the model in which the capture parameter was constrained to 0 ($G^2(18) = 1889.32$, critical = 62.99, BIC = 1946.67) nor the capture model ($G^2(17) = 1172.08$, critical = 61.67, BIC = 1238.98) provided good fit to the data. However, once again the capture model provided significantly better fit to the data than did the model that did not include the capture parameter ($\Delta G^2(1) = 717.24$, critical = 36.08). This, along with the lower BIC value in the capture model than in the model without capture, suggests that inhibitory control – or another cognitive ability resulting the same pattern of responses – influenced responding in the SPIN task. Thus, as was true for false memory, results of the MPT modelling contradicted the findings of the auditory Stroop task, suggesting that inhibitory control may have influenced susceptibility to false hearing. Reasons for these contradictory findings are discussed below.

I again plotted the observed and predicted values for the capture model to determine whether its fit to the SPIN data was as poor as the G^2 value suggested (see Figure 13). The plot suggests that the basic capture model struggled to fit correct responses rated as hear or know, "other" responses (i.e., responses other than the target word or the pre-selected alternative to the target word), and false hearing responses rated as hear or know in the incongruent condition, but generally did well with other response types. However, since younger and older adults differed in their responding in the SPIN task, we would expect better fit for models that account for age differences.

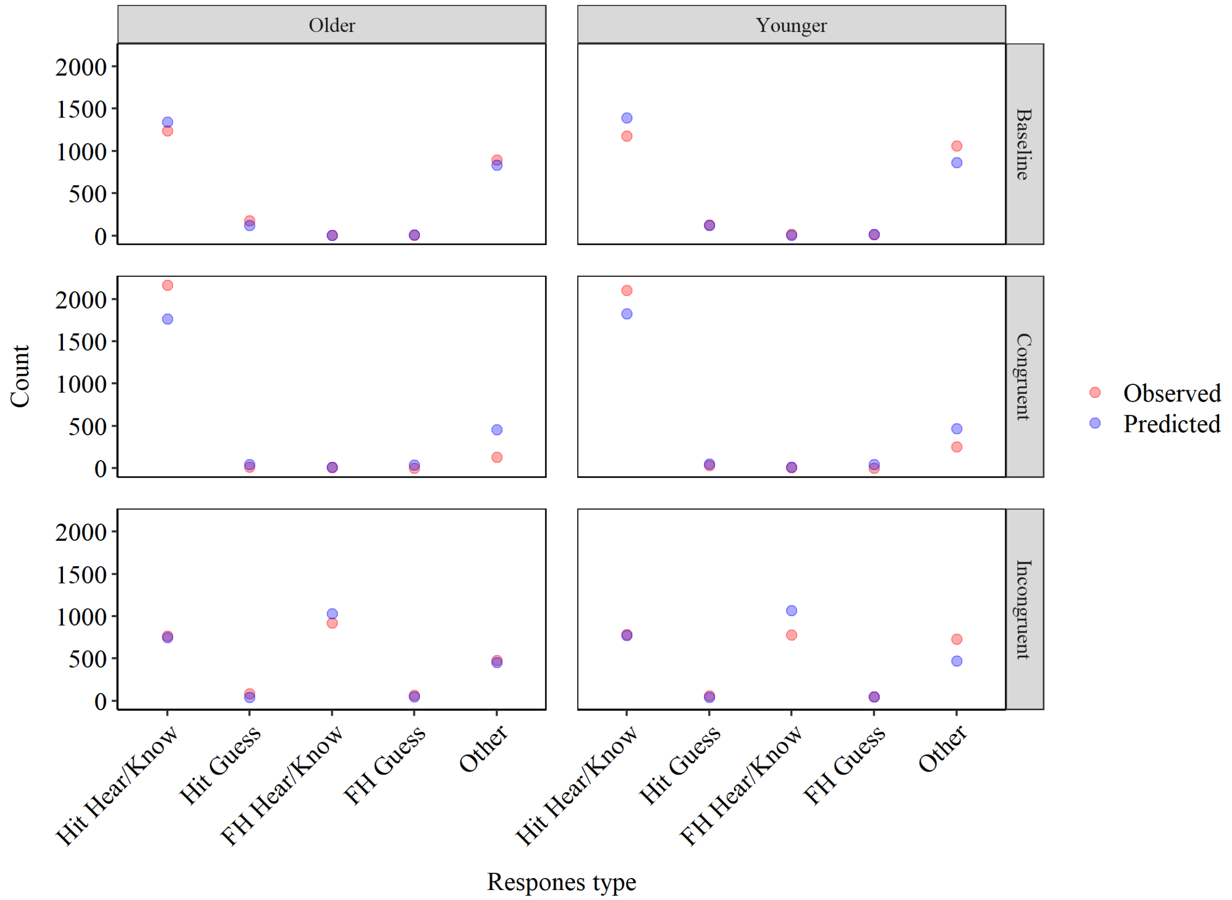

Figure 13. Observed versus predicted counts of responses in each context condition for the base capture model of hearing. FH = false hearing; Other = responses other than the target word or the pre-selected alternative word.

I next created models in which each individual parameter was allowed to vary between age groups while the other parameters were held constant and compared these models to the basic capture model in which all parameters were held constant to determine whether age groups differed in any of the parameters. I found age differences in capture ($\Delta G^2(1) = 46.66$, critical = 36.08, BIC = 1201.88), hearing ($\Delta G^2(1) = 41.53$, BIC = 1207.01), and word generation (ΔG2(1) = 37.45, BIC = 1220.65), but not in accessibility bias ($\Delta G^2(1) = 33.20$, BIC = 1224.91) or attribution threshold ($\Delta G^2(1) = 13.02$, BIC = 1235.53). Looking at the individual parameter estimates revealed that older adults were more likely to be captured by context than younger adults (.49 vs. .39), were more likely to hear the word than younger adults (.60 vs. .53), and were less likely to generate the pre-selected alternative word in the baseline condition but more likely to do so in the congruent and incongruent conditions than younger adults (in baseline condition: .01 vs. .03; in congruent/incongruent conditions: .16 vs. .07). Again, while it may seem odd that older adults were classified as more likely to hear the target word than younger adults given past studies showing age-related hearing loss (Morrell et al., 1996; Sommers et al., 2011), this difference simply reflects older adults' higher accuracy in the baseline condition of the SPIN task resulting because the SNRs chosen for each age group failed to equate performance.

Finally, I created models allowing all possible combinations of parameters to differ across younger and older adults. Table 6 lists the G^2 and BIC values for each of these models. Comparing these models based on their BIC values, the best fitting model allowed for group differences in the capture, hearing, word generation, and attribution threshold parameters. This model was still classified as a poor fit to the data according to its G^2 value ($G^2(12) = 1042.00$, critical = 54.85). Additionally, as can be seen in Figure 14, the model still had trouble fitting correct responses rated as hear or know in the baseline and congruent conditions, "other" responses, and false hearing responses rated as hear or know in the incongruent condition. This suggests that there may be additional parameters beyond those considered here that are needed to explain responding in the SPIN task.

Table 6

Fit statistics for each MPT model in the SPIN task

Parameters allowed to vary	G^2	BIC
None	1172.08	1238.98
C	1125.42	1201.88
H	1130.55	1207.01
A	1138.88	1224.91
W	1134.63	1220.65
AT	1159.06	1235.53
C, H	1101.83	1187.86
C, A	1096.5	1192.08
C, W	1087.97	1183.55
C, AT	1105.5	1191.52
H, A	1104.21	1199.79
H, W	1095.93	1191.51
H, AT	1099.35	1185.37
A, W	1101.83	1206.97
A, AT	1138.6	1234.18
W, AT	1127.8	1223.38
C, H, A	1077.01	1182.15
C, H, W	1066.6	1171.74
C, H, AT	1066.99	1162.57
C, A, W	1059.46	1174.16
C, A, AT	1092.22	1197.36
C, W, AT	1075.73	1180.87
H, A, W	1069.91	1184.61
H, A, AT	1092.93	1198.07
H, W, AT	1074.32	1179.46
A, W, AT	1101.08	1215.78
C, H, A, W	1042.06	1166.32

C, H, A, AT	1061.74	1176.43
C, H, W, AT	1042	1156.7
H, A, W, AT	1065.08	1189.34
C, H, A, W, AT	1034.48	1168.29

Note. BIC = Bayesian Information Criterion; C = Capture; H = Hear; A = Accessibility bias; W = Word generation; AT = Attribution threshold.

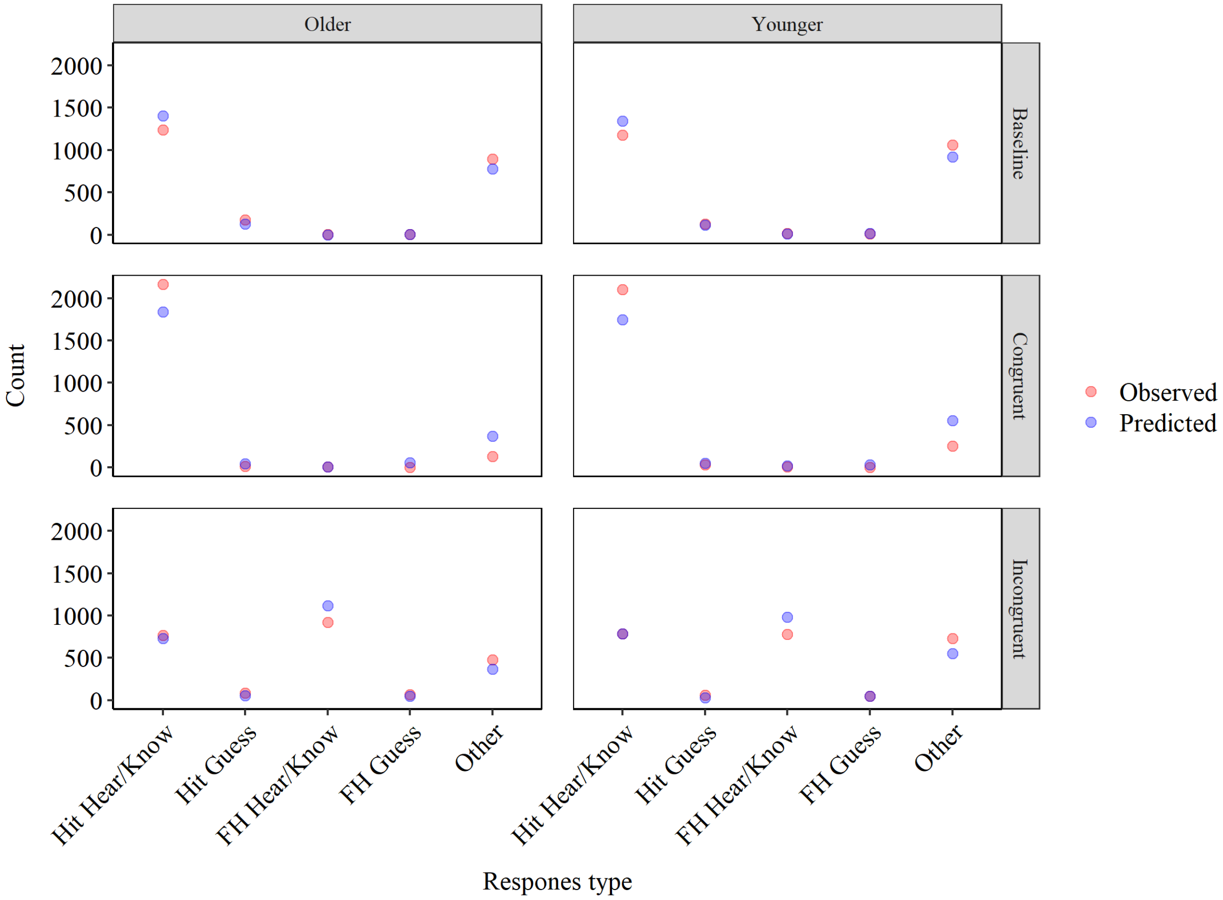

Figure 14. Observed versus predicted counts of responses in each context condition for the capture model of hearing allowing for group differences in the capture, hearing, word generation, and attribution threshold parameters. FH = false hearing; Other = responses other than the target word or the pre-selected alternative word.

2.3 Experiment 1 discussion

In Experiment 1, I successfully replicated findings of past studies of false memory (Jacoby et al., 2005, Experiment 2) and false hearing (Rogers et al., 2012; Sommers et al., 2015). I found that older adults' responses in both the memory and SPIN tasks were more likely to be based on available contextual cues than those of younger adults, as demonstrated by increased benefit from valid contextual cues and increased detriment from invalid contextual cues. Additionally, older adults were more susceptible than younger adults to both false memory and false hearing, showing that older adults were more likely to respond based on available contextual cues than younger adults even when these cues were invalid.

The primary goal of Experiment 1 was to test the hypothesis that false hearing and false memory result from similar cognitive mechanisms. In support of this claim, I found a significant positive relationship between susceptibility to false hearing and susceptibility to false memory. This was the first demonstration of a direct relationship between these two phenomena. I then investigated the roles of two potential mechanisms that may underlie this relationship, each of which has been shown to be predictive of false memory in past studies: inhibitory control and executive functioning.

Results with regards to inhibitory control were mixed. Inhibitory control as measured by the auditory Stroop task was not a significant predictor of either false memory or false hearing. This was surprising given that a relationship between inhibitory control and false memory was observed by Sommers and Huff (2003, Experiment 2) using the same auditory Stroop task.

There are several potential explanations for why the predicted relationship between inhibitory control as measured by the auditory Stroop task and false hearing and false memory did not emerge in the present study. For example, it is possible that the auditory Stroop task was not sensitive enough to pick up on age differences in inhibitory control. Whereas both younger and older adults experienced the typical Stroop effect (i.e., slower response time in the incongruent condition than in the baseline condition), there was no difference in the magnitude of the Stroop effect across groups, suggesting that younger and older adults did not differ in the response suppression aspect of inhibitory control. Since I observed significant age differences in susceptibility to both false hearing and false memory in the present study, the lack of age differences in the magnitude of the Stroop effect would mitigate the correlation between auditory Stroop performance and both false hearing and false memory. It is possible that the auditory Stroop task used in the present study accurately measured individual differences in the response suppression aspect of inhibitory control, in which case we would conclude that younger and older adults did not differ in this ability and that this ability was unrelated to susceptibility to false hearing and false memory. However, the lack of age differences in the Stroop effect in the present study was surprising given the extensive literature demonstrating an age-related decline in inhibitory control (Cohn et al., 1984; Hasher & Zacks, 1988; Hasher et al., 1997; Jacoby et al., 2005; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West & Alain, 2000) and the finding of increased interference in older, relative to younger, adults on an identical auditory Stroop task in the study by Sommers and Huff (2003, Experiment 2). The auditory Stroop paradigm has demonstrated inconsistency in detecting age differences in the response suppression aspect of inhibitory control. Whereas some studies using the auditory Stroop paradigm have found that older adults demonstrated greater interference than younger adults (Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2), two recent studies detected no age differences (Mazaheri et al., 2015; Vervoort et al., 2019)[8]. The inconsistency of the auditory Stroop paradigm with regards to detecting age differences in inhibitory control raises the concern that the null age difference in the present study reflected lack of measurement sensitivity or reliability as opposed to lack of age differences in inhibitory control per se. Therefore, it could be the case that the ability to suppress a prepotent, but incorrect, response contributed to susceptibility to false hearing and false memory, but this ability was not accurately measured by the auditory Stroop task.

A second possible reason for the discrepancy in findings with regards to age differences in inhibitory control between the present study and the study by Sommers and Huff (2003, Experiment 2) was that the statistical methods used to analyze the data differed across the two studies. I used mixed-effects linear regression to analyze the Stroop data and mixed-effects logistic regression to determine the relationship between Stroop reaction times and susceptibility to trial-by-trial false hearing and false memory, whereas Sommers and Huff used repeated-measures ANOVA to analyze their Stroop data and linear regression to determine the relationship between Stroop performance and average susceptibility to false memory. Mixed-effects modeling has the benefit of accounting for within-subject variability by using individual trial outcomes as opposed to averages, but it is possible that accounting for this variability suppressed the effect observed by Sommers and Huff. To test this possibility, I re-ran my analyses using the same statistical method as Sommers and Huff. As in my original analyses, I

---

[8] In the studies by Mazaheri et al. (2015) and Vervoort et al. (2019), the auditory Stroop task involved listening to the words *high* and *low* spoken in high- or low-pitched voices and identifying the pitch of the voice, whereas the task in the studies by Sommers and Danielson (1999) and Sommers and Huff (2003) was to identify the gender of the speaker (male/female) saying the words *male* (*father* in Sommers and Danielson), *female* (*mother*), or *person*. Additionally, participants in the study by Mazaheri et al. (2015) completed the task while walking on a treadmill, whereas participants in the other studies completed the task while seated. It is possible that these methodological differences resulted in the different findings across studies.

found that there was a significant Stroop effect using repeated measures ANOVA, $F(1,114) = 163.71$, $p < .001$. However, there was once again no difference across age groups in the magnitude of the Stroop effect, $F(1,114) = .87$, $p > .05$. Similarly, adding incongruent Stroop reaction time to a linear regression model with age group and average baseline Stroop reaction time did not improve model fit when predicting average susceptibility to either false memory, $F(1,114) = 1.20$, $p > .05$, or false hearing, $F(1,114) = 2.48$, $p > .05$, suggesting that the response suppression aspect of inhibitory control, as measured by the auditory Stroop task, was unrelated to these phenomena. Thus, differences in statistical methods cannot account for the failure to replicate Sommers and Huff's findings in the present study.

Another difference between the present study and the study by Sommers and Huff (2003, Experiment 2) is the memory task that was used. I used the process dissociation task from Jacoby et al. (2005, Experiment 2), whereas Sommers and Huff used a phonological DRM paradigm. It is possible that different paradigms used to evoke false memories place different requirements on inhibitory control, and the process dissociation task I used required less inhibitory control, or a different aspect of inhibitory control, than a phonological DRM paradigm. There is evidence showing that susceptibility to false memories in the DRM paradigm is uncorrelated with susceptibility to false memories in the misinformation paradigm (Ost et al., 2013; Patihis et al., 2018) and that there is no significant correlation between false memories evoked using semantic DRM paradigms and phonological DRM paradigms (Ballou & Sommers, 2008). Therefore, it is plausible that false memories evoked using different paradigms rely on different cognitive mechanisms. However, this would not account for the lack of age differences in the magnitude of the Stroop effect. The inconsistency of age differences in the Stroop effect between the present study and the study by Sommers and Huff makes it impossible to evaluate whether the failure to replicate their findings resulted from differences in the Stroop task, the memory task, or elsewhere.

The results of the EFA showing that the Stroop difference score did not load onto the same latent factor as the measures of executive functioning raises further concerns regarding the auditory Stroop task as a measure of inhibitory control. Since inhibitory control is typically considered a component of executive functioning (Jurado & Rosselli, 2007), we would expect it to load onto the same factor as other measures of executive functioning. Showing that this was not the case raises uncertainty regarding what the auditory Stroop task measured. It could be the case that auditory Stroop performance indexed an aspect of inhibitory control that was not captured by other measures of executive functioning in the present study, which would explain why these measures did not load onto the same factor. In their update of Hasher and Zacks' (1988) original theory of inhibitory control, Lustig et al. (2007) detailed three distinct functions of inhibitory control: stopping irrelevant information from entering working memory, removing irrelevant information from working memory, and suppressing highly activated, but incorrect, responses. Therefore, it is possible that the auditory Stroop task tapped into a different component of inhibitory control than the measures included in the executive functioning battery. For example, the role of inhibitory control in the Wisconsin Card Sorting Test may be to remove the previous sorting rule from working memory when it is no longer relevant, whereas the role of inhibitory control in the auditory Stroop task is thought to be to suppress the response activated by the spoken word. This conclusion is supported by past work showing that the number of categories achieved on the Wisconsin Card Sorting Test loaded onto a different factor than Stroop interference (Boone et al., 1998; Pineda & Merchan, 2003; Rodríguez-Aranda & Sundet, 2006). However, Rodríguez-Aranda and Sundet (2006) also found that Stroop interference loaded onto the same factor as the number of words generated on the COWAT, another of the measures included in the executive functioning battery. I found no correlation between Stroop interference and the number of words generated on the COWAT in the present study, and found that these variables loaded onto different factors in the EFA. Given that relationships between Stroop interference and other measures of executive functioning seem to be inconsistent or absent, this suggests that the auditory Stroop task measured something that was not captured by the executive functioning battery. It is possible, for example, that differences in reaction time between the baseline and incongruent conditions of the auditory Stroop task indexed something other than inhibitory control, like general confusion or strategic slowing to avoid errors. Due to the uncertainty surrounding the sensitivity and validity of the auditory Stroop task as a measure of inhibitory control in the present study, researchers should be cautious when interpreting findings involving the auditory Stroop task.

Although results involving the auditory Stroop task indicated that the response suppression aspect of inhibitory control might be unrelated to false hearing or false memory, evidence for the role of inhibitory control in these phenomena was provided by the MPT models. Specifically, adding the capture parameter – interpreted by Jacoby et al. (2005) as representing inhibitory control – improved the fit of MPT models for both the SPIN task and the memory task. Additionally, as in the study by Jacoby et al. (2005, Experiment 2; see also Millar et al., 2018 for more recent use of the capture model), the best-fitting MPT model to the memory data suggested that there were significant age differences in susceptibility to capture, supporting age differences in inhibitory control observed in past studies (Cohn et al., 1984; Hasher & Zacks, 1988; Hasher et al., 1997; Jacoby et al., 2005; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West & Alain, 2000). Therefore, Experiment 1 provided some evidence for the role of inhibitory control in false hearing and false memory, and suggested that age differences in at least one aspect of inhibitory control may account, in part, for age differences in susceptibility to false memory. However, as described above, it is possible that the capture parameter may represent processes other than, or in addition to, inhibitory control, and thus this finding should be interpreted with caution.

The MPT models also suggested a critical difference between responding in the memory task and responding in the SPIN task. Whereas age differences in the memory task could be accounted for by differences in the capture and recollection parameters, the best fitting model for the SPIN data allowed age differences in the capture, hearing, word generation, and attribution threshold parameters (i.e., all but the accessibility bias parameter). Even with this far more complex model, the capture model was still a poor fit to the SPIN data, whereas a much simpler model provided relatively good fit to the memory data. It is clear that additional parameters would be necessary to explain responding in the SPIN task. For example, given that hits characterized as hear/know tended to be overestimated in the baseline condition but underestimated in the congruent condition, specifying a separate hearing parameter in the baseline condition may help to resolve this issue. This change is theoretically justifiable given that participants may need to try harder to hear the target word in the absence of contextual cues, whereas the contextual cues may clue participants in to what sounds they should listen for in the congruent and incongruent conditions. Similarly, specifying a separate accessibility bias parameter in the congruent condition may help to resolve the tendency of the model to underestimate “other” responses in the baseline and incongruent conditions and overestimate “other” responses in the congruent condition. This change also makes sense given that we would expect the target word to be particularly accessible in the congruent condition due to the availability of valid contextual cues.

To test this hypothesis, I re-ran the MPT model for the SPIN task specifying these additional parameters. The basic capture model with no group differences between parameters was still classified as a poor fit to the data ($G^2(15) = 662.17$, critical = 58.99, BIC = 748.19), but the predicted counts showed much closer correspondence to the observed counts than the models above that did not specify condition differences in hearing and accessibility bias (see Figure 15). Interestingly, the best-fitting model included group differences in only two parameters: accessibility bias and word generation. Older adults were more biased towards the most easily accessible word in all conditions (baseline = .58, congruent = .85, incongruent = .32) than were younger adults (baseline = .50, congruent = .73, incongruent = .18). Older adults were also more likely to generate the alternative word in the congruent and incongruent conditions than younger adults (.30 vs. .16), but the likelihood of generating the alternative word in the baseline condition was approximately equivalent across groups (.01 vs. .03)[9]. This MPT model was also classified as a poor fit to the data ($G^2(10) = 519.96$, critical = 51.97, BIC = 653.78), although there was very good correspondence between the predicted and observed counts (see Figure 16).

---

[9] The likelihood of generating the alternative word in the baseline condition was much lower than in the combined congruent/incongruent conditions because the alternative word in the incongruent condition was the contextually predicted word (e.g., *box* for the sentence *She put the toys in the fox*). In the baseline condition, the alternative word was a phonological neighbor of the target word selected at the outset of the study, and there were no contextual cues from which to predict this word (e.g., *cage* for the sentence *The word is page*). Thus, the high predictability of the alternative word in the incongruent condition resulted in the higher word generation parameter in the combined congruent/incongruent conditions relative to the baseline condition.

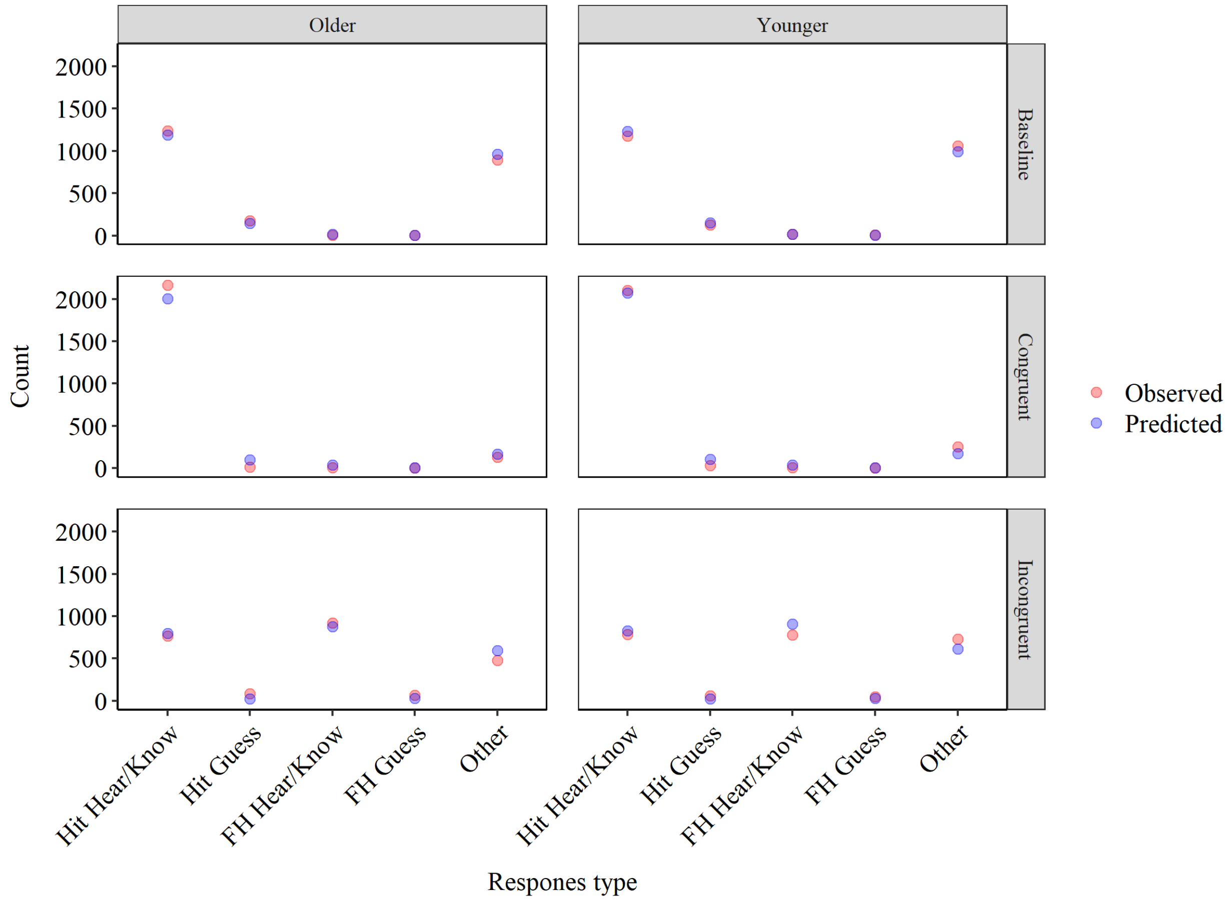

Figure 15. Observed versus predicted counts of responses in each context condition for the basic capture model of hearing specifying a different hearing parameter in the baseline condition and a different accessibility bias parameter in the congruent condition. FH = false hearing; Other = responses other than the target word or the pre-selected alternative word.

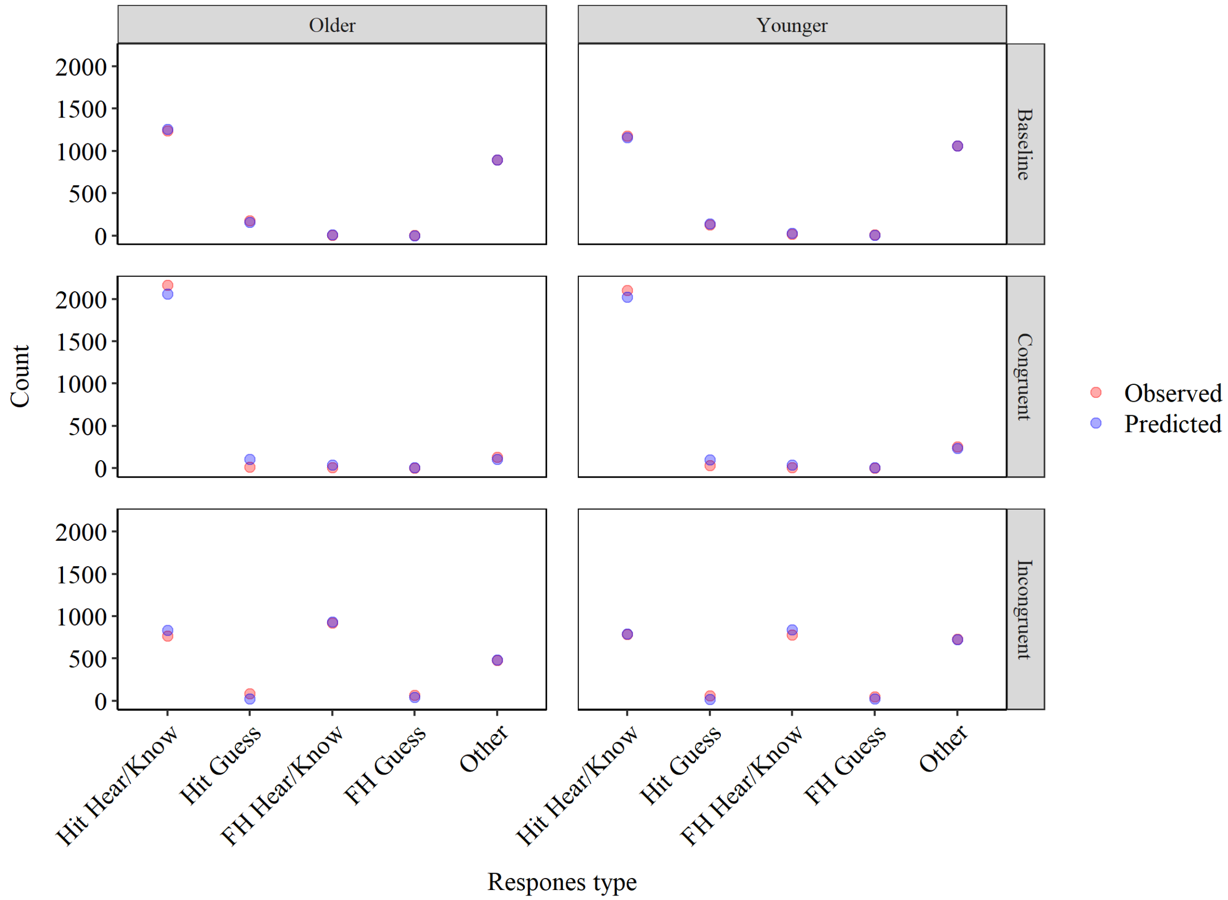

Figure 16. Observed versus predicted counts of responses in each context condition for the capture model of hearing specifying a different hearing parameter in the baseline condition and a different accessibility bias parameter in the congruent condition allowing age differences in accessibility bias and word generation. FH = false hearing; Other = responses other than the target word or the pre-selected alternative word.

When I re-ran the MPT analysis on the memory data with the same parameter changes, I found that the best fitting model was still one in which age differences were permitted in capture and recollection ($G^2(12) = 60.66$, BIC = 171.90). Additionally, there was no significant difference in fit between the original best-fitting model presented in the Results section above and the more complex model described here ($\Delta G^2(1) = 3.25$, critical = 27.22), suggesting that the additional parameters needed to fit the SPIN data were not necessary to fit the memory data. Together with the findings above, this suggests that the parameters specified in the capture model may be able to account for responding in both the memory and SPIN tasks, but increased model specificity is needed to accurately model responding in the SPIN task. Additionally, my findings suggest that different processes may underlie age differences in responding in the memory and SPIN tasks. Whereas age differences in the memory task could be accounted for by differences in susceptibility to capture and ability to recollect the studied word, differences in the SPIN task were best accounted for by differences in bias towards the most easily accessible word and in the likelihood of generating the pre-selected alternative word. Thus, while similar cognitive processes may be at work in both the memory and SPIN tasks, different processes may underlie age differences in the two tasks.

Whereas evidence for the role of inhibitory control in false hearing and false memory was mixed, executive functioning demonstrated consistent and significant relationships with both phenomena. Specifically, participants who had better scores on the executive functioning composite tended to be less susceptible to both false memory and false hearing after controlling for age group. The relationship between executive functioning and false hearing also remained significant after accounting for individual differences in hearing acuity, lending further credence to the idea that false hearing results, in part, due to cognitive deficits as opposed to solely sensory deficits (Rogers et al., 2012; Sommers et al., 2015). Importantly, the relationship between age group and susceptibility to both false hearing and false memory became non-significant after controlling for the executive functioning battery. This suggests that differences in executive functioning may be able to fully explain older adults' increased susceptibility to false hearing and false memory observed here and in past studies (Balota et al., 1999; Jacoby, 1999; Jacoby et al., 2005; Kensinger & Schacter, 1999; Mitchell et al., 2003; Roediger & Geraci, 2007; Rogers et al., 2012; Sommers et al., 2015; Watson et al., 2001).

Executive functioning also accounted for significant variance in both false hearing and false memory after controlling for the response suppression aspect of inhibitory control as indexed by the auditory Stroop task. Whereas this may seem to suggest that executive functions other than the response suppression aspect of inhibitory control played a role in false hearing and false memory, this conclusion requires additional investigation given the results of the EFA showing that the Stroop difference score did not load on the same latent factor as the measures of executive functioning. This indicates that the mechanism causing slowing in the incongruent condition of the auditory Stroop task relative to the baseline condition – be it the response suppression aspect of inhibitory control or some other process – was best considered separate from the latent executive functioning variable. Because of this complication, we cannot draw the conclusion that controlling for Stroop performance accurately partialled out variance attributable to the response suppression aspect of inhibitory control in the executive functioning battery. Therefore, it could be the case that differences in one or more aspects of inhibitory control that remained in the executive functioning battery after controlling for aspects of inhibition assessed by the auditory Stroop task accounted for the relationships between executive functioning and false hearing and false memory. However, it could also be the case that executive functions other than inhibitory control accounted for these relationships. This will be discussed further in the General Discussion.

Another novel contribution of the present study was an examination of the subjective experience of cases of false hearing using a hear/know/guess paradigm. I found that both younger and older adults were more likely to characterize cases of false hearing as heard than either known or guessed, suggesting that cases of false hearing may be accompanied by a subjective experience akin to veridical hearing. In the memory task, however, both younger and older adults displayed relatively similar odds of characterizing false memories as remembered versus known or guessed. In fact, younger adults were numerically most likely to characterize cases of false memory as guessed and least likely to characterize cases of false memory as remembered, the opposite of the pattern observed for cases of false hearing. This suggests that there may be qualitative differences between experiencing false memory and experiencing false hearing, with false hearing being more easily confused with veridical hearing than false memory with veridical memory.

It is possible that characteristics of the memory task used in the present study gave rise to the approximately equal likelihood of characterizing false memories as remembered, known, or guessed. Past studies have shown that the relative likelihood of characterizing false memories as remembered versus known depends on the task being used, with DRM paradigms using semantically related lists resulting in a higher rate of false memories characterized as remembered than known (Norman & Schacter, 1997, Experiment 1; Roediger & McDermott, 1995, Experiment 2; Watson et al., 2003, Experiment 3), and DRM paradigms using phonologically related lists resulting in a higher rate of false memories characterized as known than remembered (Watson et al., 2003, Experiment 3). Therefore, it is possible that the high rate of false memories judged as known or guessed in the present study may result from features of the memory task itself as opposed to reflecting a common feature of all false memories. I believe that the high proportion of know and guess ratings in the memory task relative to the SPIN task resulted from it being easier to generate the contextually predicted word in this particular memory task. Specifically, the remaining letters in the fragment completion task limited the scope of plausible responses to words that could complete the fragment. For example, for the cue-fragment pair head – s--l-, there are a limited number of words that could complete the fragment, increasing the likelihood of generating the response associated with false memory (scalp) when participants were not captured by context or responding based on accessibility, but rather were simply guessing a word that could plausibly complete the fragment. This could account for the higher proportion of false memory responses being characterized as known or guessed relative to remembered. In the SPIN task, however, the target word was fully masked by background noise and there was no constraint on what word could be provided as a response. Thus, participants may have been more likely to generate a response other than either the predicted word or the target word (i.e., an "other" response) unless they were able to hear (or falsely hear) the target word in the incongruent condition. This would result in the majority of cases of false hearing being characterized as heard as opposed to known or guessed. To test the hypothesis that participants were more likely to generate an "other" response in the incongruent condition of the SPIN task than of the memory task, I conducted an ANOVA with age group, task (memory vs. SPIN), and the age group by task interaction predicting proportion of "other" responses on incongruent trials. The age group by task interaction was significant, $F(1,232) = 18.47$, $p < .001$, so I conducted t-tests within each age group to determine whether tasks differed in proportion of "other" responses. In support of my hypothesis, younger adults were more than nine times less likely to generate an "other" response in the memory task ($M = .04$) than in the SPIN task ($M = .30$), $t(106.49) = 17.50$, $p < .001$. Older adults were also less likely to generate an "other" response in the memory task ($M = .03$) than in the SPIN task ($M = .21$), $t(79.10) = 14.86$, $p < .001$, but as indicated by the significant interaction in the ANOVA, the difference between tasks was smaller than for younger adults. This supports the conclusion that participants characterized more cases of false memory as known or guessed relative to cases of false hearing because the scaffolding provided in the memory task made it easier to generate the contextually predicted word in cases where the participant could not remember (or falsely remember) the studied word. Therefore, differences observed in the subjective judgements of false hearing and false memory may reflect task differences as opposed to indicating differences in the subjective experience of these phenomena per se.

Given that inhibitory control has previously been shown to be related to false memory (Colombel et al., 2016; Sommers & Huff, 2003, Experiment 2) and is hypothesized to be a primary mechanism underlying false hearing in older adults (Rogers et al., 2012; Sommers et al., 2015), it was surprising that inhibitory control as assessed by the auditory Stroop task was unrelated to both false memory and false hearing in the present study. In Experiment 2, I further examined the potential role of inhibitory control in false hearing. First, I tested whether the predictive strength of sentences was related to the likelihood that false hearing would occur. Then I used eye-tracking to gain insight into how sentence context influenced activation of the sentence-final target word and to examine age differences in the ability to suppress this activation when context turned out to be invalid.

Chapter 3: Experiment 2

False hearing is often described as resulting from an inability to suppress an expected response, reflecting a failure of inhibitory control (Rogers et al., 2012; Sommers et al., 2015). Using the stimuli from Experiment 1 as an example, the sentence "*She put the toys in the…*" creates a strong expectation of what word should follow (*box*). In the case of incongruent sentences, the participant must then suppress their expectation of the word *box* to correctly hear the presented (but unpredicted) word *fox*. As described above, within the expanded version of the NAM (Luce & Pisoni, 1998) proposed by Sommers and Danielson (1999), increasing expectation based on context can be thought of as increasing activation of the contextually predicted word in the mental lexicon. This increased activation due to expectation increases the likelihood of accurately hearing the predicted word when context is valid (i.e., in the congruent condition) and increases the likelihood of experiencing false hearing when context is misleading (i.e., in the incongruent condition). If false hearing results from an inability to suppress an expected response, then the likelihood of false hearing may depend, in part, on the predictive strength of the preceding sentence. Specifically, in the incongruent condition, sentences that are stronger predictors of a phonological neighbor of the sentence-final word should result in higher rates of false hearing due to increased activation of the predicted (but unpresented) word relative to less predictive sentences.

Although the effect of predictive strength has yet to be investigated with regards to false hearing, it has been investigated in false memory. Indeed, in Deese’s (1959) seminal study that would later give rise to the DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), the likelihood of falsely recalling a critical word in a list-memory paradigm was shown to be positively correlated with the associative strength between the words in the studied list and the critical word (backward associative strength; see Figure 17). More recently, this effect was replicated by Roediger et al. (2001b), who sought to determine which characteristics of critical words in the DRM paradigm were related to the likelihood of false recall. The authors found that backward associative strength was the best predictor of false recall and the second-best predictor of false recognition (after only veridical recall) in the DRM paradigm. Therefore, the degree to which a critical word is predicted by the words in the studied list is positively associated with the likelihood that a false memory will occur.

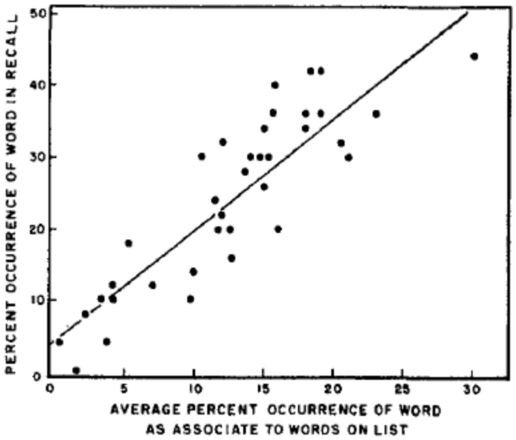

Figure 17. The correlation (*r* = .87) between the likelihood of falsely recalling a given word and the likelihood of generating that word as an associate of other words in the list from Deese (1959).

In Experiment 2, I explored the possibility that the predictive strength of sentence stimuli influenced the likelihood that false hearing would occur. I conducted analyses on data from both Experiments 1 and 2 to determine whether the cloze probability of sentences was predictive of susceptibility to false hearing. Cloze values were obtained from a sample of 34 participants on Amazon's Mechanical Turk for each sentence using a visual sentence-completion task. The sentence-final target words were removed from each sentence (e.g., She put the toys in the _____), and participants were instructed to type the first word that came to mind to complete each sentence. In line with past findings in false memory (Deese, 1959; Roediger et al., 2001b), I predicted that incongruent sentences with higher cloze values – indicating greater predictive strength – would be more likely to result in false hearing than sentences with lower cloze values. This would support the activation-based account of false hearing described above drawing upon the NAM (Luce & Pisoni, 1998) and the findings of Sommers and Danielson (1999), suggesting that stronger contextual cues convey more activation to the predicted word than weaker contextual cues, increasing the likelihood of hearing the predicted word even when it is not presented.

To supplement the cloze value analyses, I conducted a novel experiment to determine how activation of different response options changed in real time using eye-tracking in a visual world paradigm. The visual world task was similar to the SPIN task from Experiment 1, but while each sentence was played, four pictures were presented on the computer screen. The pictures depicted the target word (e.g., box), a phonological neighbor of the target word (e.g., fox), and two words that did not sound like the target word and were not predicted by the sentence context (e.g., key and paw). Using eye-tracking, I was able to determine how the proportion of fixations on each of the images changed over the course of the sentence and after the target word was presented. As mentioned above, the proportion of fixations on each image in the visual world paradigm can be used as a proxy for changes in activation, with a greater proportion of fixations relative to other images indicating higher activation (Tanenhaus et al., 2000). Based on the findings of past studies using eye-tracking (Ito et al., 2018; Kukona, 2020), hearing incongruent sentences should lead to increased fixations on an image depicting the contextually predicted (but incorrect) word before the target word is presented, reflecting increased activation of the response supported by context. If participants are able to suppress the activation of the contextually predicted word when the unpredicted target word is presented, fixations on the image depicting the contextually predicted word should decrease and fixations on the image depicting the target word should increase following presentation of the target word. The eye-tracking data also allowed me to test for age differences in the response suppression aspect of inhibitory control by comparing younger and older adults’ likelihood of decreasing their proportion of fixations on the contextually predicted image after the target word was presented in incongruent sentences.

I formed specific hypotheses regarding how younger and older adults’ fixation patterns would change over the course of baseline, congruent, and incongruent sentences. In the baseline condition, I predicted that the proportion of fixations on each of the images should remain approximately equal until the target word was presented since there were no contextual cues upon which to form an expectation. After the target word was presented in the baseline condition, fixations on the target image should increase in accordance with participants’ ability to accurately hear the target word. In congruent and incongruent sentences, I hypothesized that both younger and older adults would become increasingly fixated on the contextually predicted image leading up to presentation of the target word, demonstrating increasing anticipatory activation of the word supported by context. I predicted that this increased focus on the contextually predicted image might be greater for older than for younger adults, reflecting older adults’ increased context-based responding, as demonstrated in Experiment 1 and in previous studies (Rogers et al., 2012; Sommers et al., 2015). Whereas fixations on the target image should continue to increase for both age groups once the target word was presented in the congruent condition, I predicted that age groups would differ in their reaction to the target word’s presentation in the incongruent condition. Specifically, I predicted that younger adults would decrease their proportion of fixations on the contextually predicted image and increase their fixations on the unpredicted target image after the target word was presented, reflecting their ability to suppress the activation of the expected word. Older adults, on the other hand, were expected to maintain or even increase their fixations on the contextually predicted image after the target word was presented in incongruent sentences, reflecting an inability to suppress the activation of the expected word. This finding would support the theory that age differences in susceptibility to false hearing result, in part, from an inability to inhibit a prepotent response.

3.1 Methods

3.1.1 Participants

Participants were 23 younger adults ages 18-29 (*M* = 21.00, *SD* = 2.68) and 19 older adults ages 66-81 (*M* = 73.31, *SD* = 4.45). I had planned to collect 30 participants in each age group, but data collection had to be halted due to the COVID-19 pandemic. All participants completed the demographics form described in Experiment 1 and the Shipley Vocabulary Test (Shipley, 1940). Younger (*M* = 33.04, *SD* = 3.04) and older adults (*M* = 34.42, *SD* = 3.91) displayed equivalent vocabulary knowledge, $t(33.61) = -1.26$, $p > .05$. As in Experiment 1, hearing thresholds were assessed for octave frequencies from 250 to 8000 Hz in a sound-attenuating booth using standard audiometry. Older adults (*M* = 23.68, *SD* = 11.78) had poorer best-ear PTAs (500/1000/2000 Hz frequencies) than younger adults (*M* = 4.42, *SD* = 3.39), $t(20.47) = -6.90$, $p < .001$. All participants were native English speakers who did not require the use of a hearing aid and self-reported normal or corrected-to-normal vision.

3.1.2 Materials

Materials were baseline, congruent, and incongruent sentences from the SPIN task in Experiment 1 in which the target words could easily be depicted in picture form. Periods of silence of different lengths were inserted at the start of each sentence so that the onset of the target word began at the same time on each trial to facilitate eye-tracking analyses. Four images were gathered for each sentence: one depicting the target word (e.g., *box* for the sentence *She put the toys in the box*), one depicting a semantically unrelated phonological neighbor of the target word that acted as the target word in incongruent sentences (alternative image; e.g., *fox*), and two semantically unrelated foil words that did not sound like the target word (e.g., *key* and *paw*)[10]. The length, frequency, phonological neighborhood density, and concreteness of target words were collected from the English Lexicon Project (Balota et al., 2007), and averages across sentence conditions are presented in Table 7. Target words did not differ significantly in terms of any of these lexical characteristics across sentence conditions, all *p*s > .05. Additionally, the alternative word and foil words were matched with the target word based on these lexical characteristics within each sentence type, *p*s > .05.

Table 7
Means and standard deviations of lexical characteristics of target words across conditions

	Baseline		Congruent		Incongruent	
	M	*SD*	*M*	*SD*	*M*	*SD*
Length	4.20	.76	4.37	.85	4.23	1.01
Frequency	33,266.53	77,434.77	28,985.50	44,531.67	32,123.90	44,336.55
Phono N	19.07	9.45	18.67	9.89	20.80	11.16
Concreteness	4.66	.26	4.71	.41	4.61	.39

Note. Phono N = Phonological neighborhood density

I selected an image depicting each target word, alternative word, and foil word, and resized each image to 300 x 300 pixels. For images that did not have equal width and height, a white border was added to the shorter dimension to achieve the 300 x 300 size. I conducted a pilot test to ensure that all images to be used in the visual world task were identifiable as the words they were meant to depict. Twenty younger adults participated in this pilot study. First,

---

[10] Two foil images were included on each trial as opposed to including a second phonological neighbor or a semantic competitor of the target word so that participants would not divide their fixations across two images tapping the same source of information (semantic or phonological). This allowed for clearer assessment of how fixations were impacted by semantic congruency and phonological similarity.

participants completed a study phase in which they saw each image along with the word the image was meant to depict for 2000 ms. Participants then completed a test phase wherein each written word was presented at the center of the screen one at a time along with four images. One of the images depicted the written word at the center of the screen, and the other three images were the images paired with the initial image on trials in the Experiment 2 SPIN task. For example, when the target word was fox, the four images on screen depicted a fox, a box, a key, and a paw, and the same four images were presented when the target word was box, key, and paw. The images were randomly assigned to one of the four quadrants of the screen. Participants were tasked with clicking on the image that depicted the word at the center of the screen. Average accuracy for identifying the image depicting each target word was 99.42%. In fact, only two images were identified correctly in less than 90% of cases, one of which was only used in the practice trials of the SPIN task (*joker*, Accuracy = 75%) and the other was a foil (*till*, Accuracy = 70%). Given the high identification accuracy of virtually all images in the pilot study and the similarity of the pilot study's procedure to that of the SPIN task in Experiment 2[11], I felt confident that participants would associate each image with the word they were intended to depict in the SPIN task.

Initial pilot testing revealed that accuracy on the SPIN task was at ceiling in the baseline condition for younger adults using the -4 dB SPL SNR from Experiment 1 (*M* = .93), and older adults' performance at the +1 dB SPL SNR from Experiment 1 was also higher than the desired .50 (*M* = .73). It was important to ensure that baseline performance was not too high so that participants had room to improve with the addition of congruent context and so that there was

---

[11] The only difference between the procedure of the pilot study and the SPIN task in Experiment 2 was that target words were visually presented at the middle of the screen in the pilot study whereas target words were presented aurally in noise at the end of sentences in the SPIN task.

enough ambiguity as to the identity of the target word for false hearing to occur. The high accuracy in Experiment 2 was unsurprising given that the SPIN task was a four-alternative forced-choice test as opposed to the open-set response format of Experiment 1. Further pilot testing was conducted to determine what SNRs should be used to equate performance in younger and older adults. It was determined that an SNR of -10 dB SPL for younger adults and -7 dB SPL for older adults would achieve approximately equal performance across groups with accuracy that left room for improvement from the baseline condition to the congruent condition.

3.1.3 Procedure

Participants first completed the demographic questionnaire and the audiogram inside a sound-attenuating booth. Then participants were seated at an EyeLink 1000 eye-tracking-enabled computer, where they completed the Shipley Vocabulary Test (Shipley, 1940) before beginning the visual world task. Participants placed their chins on a chinrest with their foreheads against a forehead rest to complete the visual world task. The distance from the back of the forehead rest to the eyepiece of the eye-tracker was 52.07 cm and the distance from the back of the forehead rest to the center of the computer monitor was 57.78 cm. Participants first completed a study phase wherein each image to be shown in the visual world task was shown with the word the image was meant to depict for 2000 ms to ensure that participants knew what each image represented. Participants then completed three practice trials, followed by 90 test trials, equally divided between the three sentence types (baseline/congruent/incongruent). Each trial began when the participant clicked on a central fixation cross. On each trial, a sentence was played through headphones with the final word in noise, and the four images associated with that sentence were presented on screen, one randomly assigned to each quadrant (see Figure 18). Participants were instructed to look at and click on the image corresponding to the word presented in noise, and were told that they could move their eyes freely about the screen as long as the images were displayed. Images remained on screen until the participant clicked on one of them. After clicking on an image, participants clicked on a number from one to five to indicate their confidence that they had selected the correct image, where one indicated a complete guess and five indicated absolute certainty.

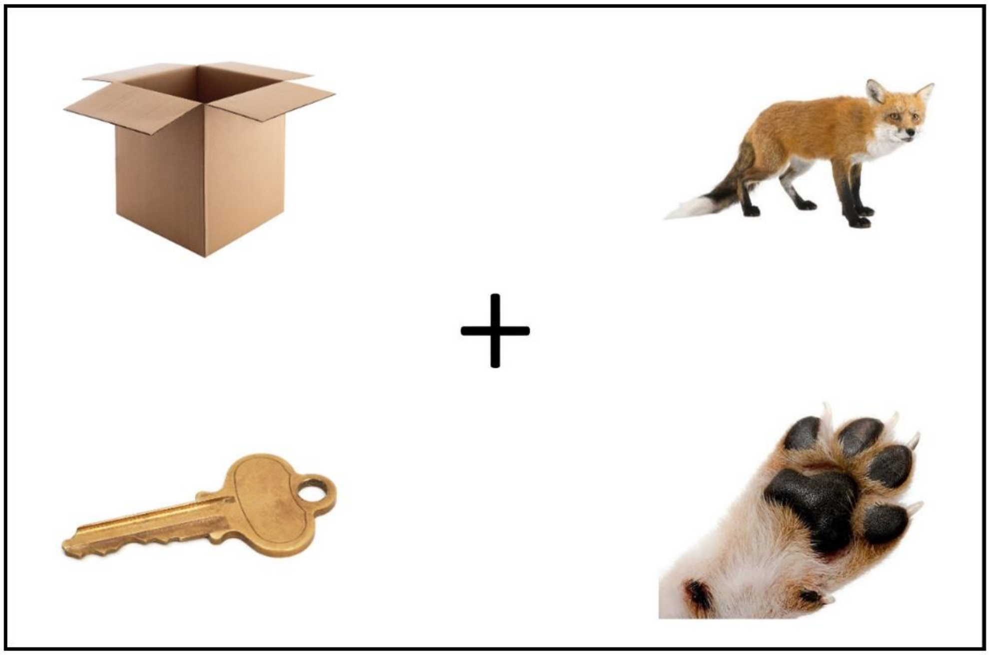

Figure 18. An example of the screen during the visual world task for the congruent sentence *She put the toys in the box*.

The eye-tracker was calibrated immediately before test trials to ensure accurate eye-tracking. For the calibration task, dots appeared at 13 different locations on the screen and participants were tasked with fixating on each dot when it appeared and continuing to look at the dot until it disappeared. Participants completed the calibration task until it had been rated a “good” calibration by the EyeLink program, then completed an additional validation calibration to ensure that calibration was consistently accurate.

3.2 Results

3.2.1 Accuracy

I first analyzed age differences in accuracy across the baseline, congruent, and incongruent conditions using the same analyses as in Experiment 1. Average accuracy in the baseline, congruent, and incongruent conditions is presented in Figure 19. Younger and older adults differed significantly in their accuracy in the baseline condition, but unlike in Experiment 1, younger adults displayed better baseline accuracy than older adults (*OR* = .64, $z = -2.05$, $p < .05$). As expected, younger adults displayed improved performance in the congruent condition (*OR* = 3.17, $z = 2.67$, $p < .01$) and poorer performance in the incongruent condition (*OR* = .29, $z = -3.25$, $p < .01$) relative to the baseline condition. As in Experiment 1, however, older adults experienced significantly greater benefit in the congruent condition (*OR* = 6.62, $z = 4.15$, $p < .001$) and significantly greater detriment to performance in the incongruent condition (*OR* = .43, $z = -2.63$, $p < .01$) relative to baseline than did younger adults. These results replicate those from Experiment 1 and support the argument that older adults’ performance was influenced more by available contextual cues than was that of younger adults.

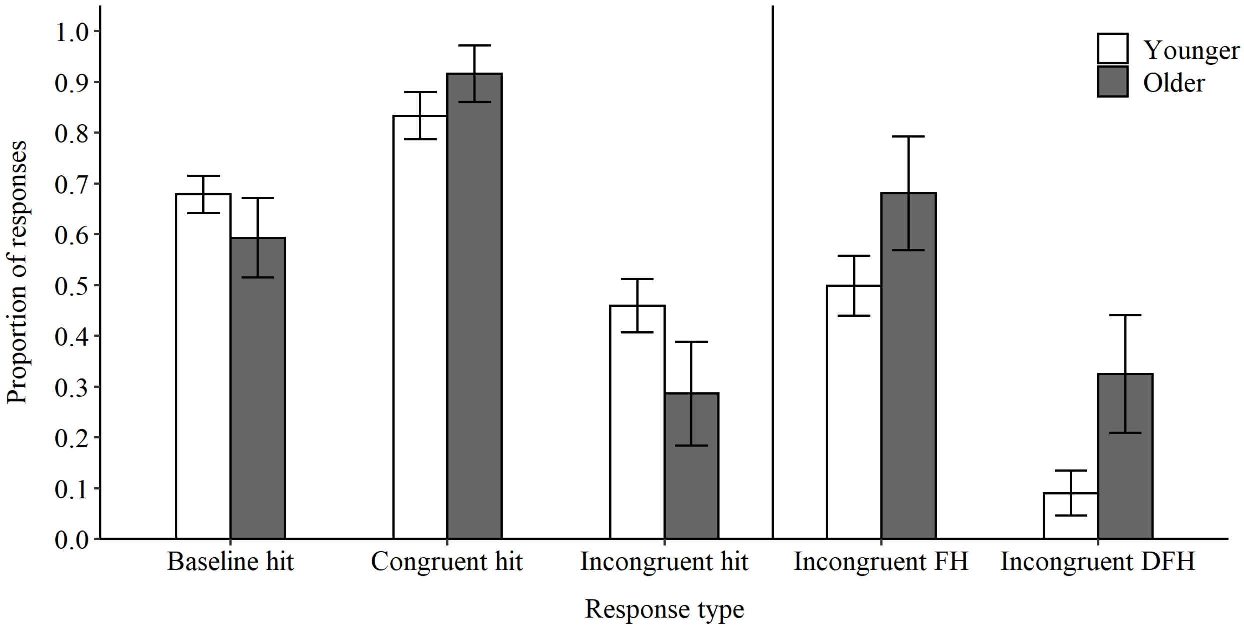

Figure 19. Average proportion of hits and susceptibility to false hearing (FH) and dramatic false hearing (DFH) for younger and older adults. Error bars represent 95% confidence intervals.

3.2.2 False hearing

I next employed the same statistical analysis as in Experiment 1 to determine whether younger and older adults differed in susceptibility to false hearing. As shown in Figure 19, both younger and older adults experienced false hearing. The odds that a younger adult would experience false hearing on incongruent trials was equivalent to the odds of giving any other response combined (*OR* = .98, z = -.05, p > .05). Older adults were more than four times as likely to experience false hearing than were younger adults, which was significant (*OR* = 4.04, z = 3.47, p < .001).

I also analyzed the odds of experiencing false hearing with maximum confidence – referred to in past studies as *dramatic false hearing* (Rogers et al., 2012; Sommers et al., 2015) – in younger and older adults (see Figure 19). The odds that a younger adult would experience dramatic false hearing was far less than the odds of giving any other response in the incongruent condition combined (*OR* = .03, z = -8.14, p < .001). Similar to previous studies (Rogers et al., 2012; Sommers et al., 2015), the odds that older adults would experience dramatic false hearing was more than 10 times greater than for younger adults (*OR* = 10.46, *z* = 4.97, *p* < .001). Thus, despite having an image depicting the correct target word presented on screen, both younger and older adults incorrectly reported hearing the contextually predicted word in over 50% of incongruent sentences, with older adults doing so more often and being far more likely to report maximum confidence in these errors than younger adults.

3.2.3 Confidence

3.2.3.1 Accurate responses

To determine whether sentence condition and age group differences existed for confidence in accurate responses, I created a linear mixed-effects regression model using the lmer function from the *lme4* package in R (Bates et al., 2015) using a subset of data that included only accurate responses. As in the accuracy analyses, the model included an intercept term corresponding to confidence in accurate responses for younger adults in the baseline condition, two dummy coded variables indicating the change in confidence from the baseline condition to the congruent and incongruent conditions for younger adults, a group variable representing the change in confidence from younger to older adults in the baseline condition, and the interaction of group with the congruent and incongruent condition dummy codes to determine whether the change in confidence from the baseline condition to the congruent and incongruent conditions differed between younger and older adults. Confidence in the baseline condition and changes in confidence from the baseline condition to the congruent and incongruent conditions were allowed to vary randomly across subjects, and confidence was allowed to vary randomly across items.

Average confidence in accurate responses (hits) in the baseline, congruent, and incongruent conditions is presented in Figure 20. Younger adults expressed confidence slightly above a neutral rating for accurate responses in the baseline condition, with the model estimating an average confidence of 3.53 out of 5. Younger and older adults' confidence did not differ in the baseline condition ($ED = .06$, $t = .32$, $p > .05$). Younger adults' confidence did not differ in either the congruent condition ($ED = .10$, $t = .43$, $p > .05$) or the incongruent condition ($ED = -.23$, $t = -1.10$, $p > .05$) relative to the baseline condition. Additionally, the difference in confidence between the baseline and incongruent conditions did not differ in older adults relative to younger adults ($ED = -.02$, $t = -.18$, $p > .05$). However, there was a significant interaction suggesting that older adults' confidence increased to a greater degree than younger adults' from the baseline condition to the congruent condition ($ED = .39$, $t = 2.18$, $p < .05$). Overall, these findings suggest that participants' confidence in accurate responses remained quite stable regardless of the context condition, aside from higher confidence in the congruent condition by older, relative to younger, adults. This differs from in past studies (Rogers et al., 2012; Sommers et al., 2015), where both younger and older adults have demonstrated lower confidence in accurate responses in the baseline condition than in the congruent condition. It is possible that changing to a four-alternative forced-choice paradigm resulted in the consistently high confidence in accurate responses in the present study. For example, if a participant thought they heard the word *box,* they might be more confident in that response because an image of a box was among the four options on screen. Therefore, the feedback provided by images on screen may have increased confidence in accurate perceptions.

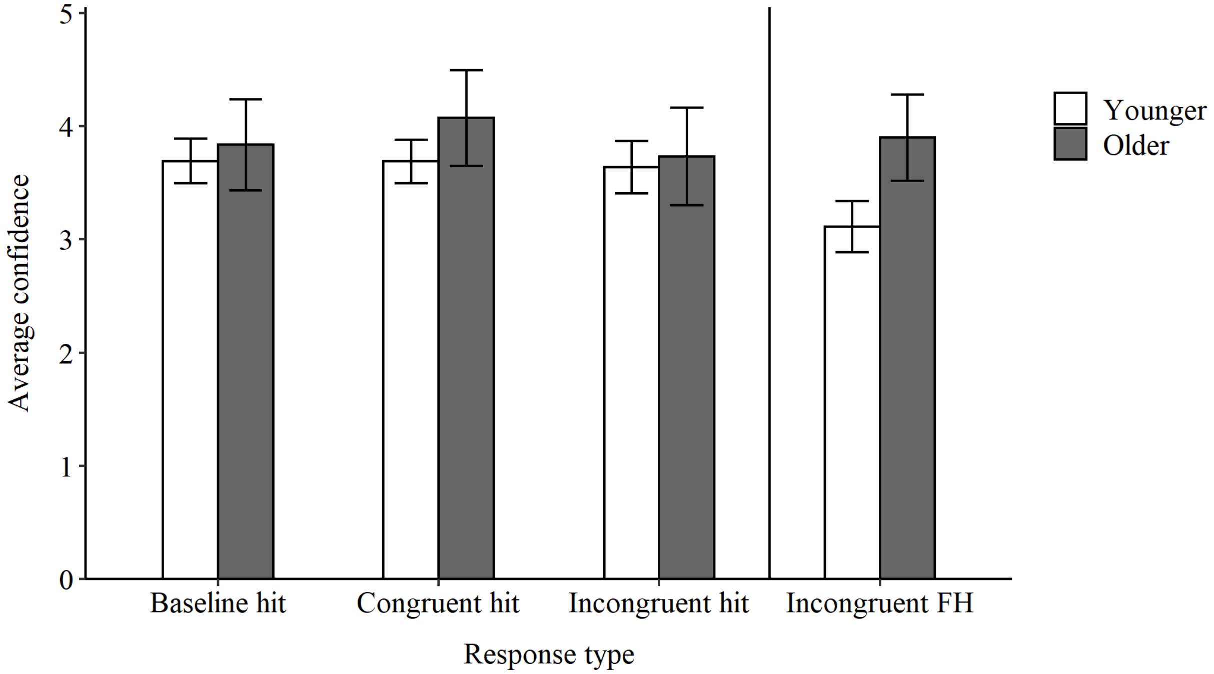

Figure 20. Average confidence in hits in the baseline, congruent, and incongruent conditions, and in cases of false hearing (FH) in the incongruent condition for younger and older adults. Error bars represent 95% confidence intervals.

3.2.3.2 False hearing responses

I then conducted a second linear mixed-effects regression analysis to determine whether younger and older adults differed in their confidence in cases of false hearing. This model was conducted on a subset of data that included only cases of false hearing on incongruent trials. The model included an intercept term corresponding to younger adults’ confidence in cases of false hearing and a dummy-coded group variable indicating the change in confidence from younger to older adults. Confidence in cases of false hearing was allowed to vary randomly across subjects and items.

Although there was little difference between age groups for confidence in accurate responses, younger and older adults did differ in their confidence in cases of false hearing (see Figure 20). Younger adults expressed approximately neutral confidence in cases of false hearing, with the model estimating an average confidence of 3.12 out of 5. Older adults' estimated average confidence in cases of false hearing was .81 higher than that of younger adults, which was a significant difference (*t* = 4.08, *p* < .001). Thus, older adults were both more susceptible to and more confident in cases of false hearing than were younger adults.

3.2.4 Cloze value analyses

The analysis of the effect of cloze value on susceptibility to false hearing was conducted with mixed-effects logistic regression using the glmer function from the *lme4* package in R (Bates et al., 2015). A similar analysis could have been conducted to examine the relationship between cloze value and the odds of an accurate response in the congruent condition, but since participants displayed nearly perfect accuracy in the congruent condition in both experiments, it was unlikely that there would be sufficient variability for a relationship to emerge. For this reason, I focused on the relationship between cloze value and susceptibility to false hearing. Since I was specifically interested in false hearing, this analysis was conducted on a subset of data that included only incongruent trials. The predictors of trial-by-trial false hearing in the model were cloze value standardized across incongruent sentences, age group, and the cloze value by age group interaction. Susceptibility to false hearing and the effect of cloze value were allowed to vary randomly across participants. Susceptibility to false hearing was not allowed to vary randomly across items because this variance may be systematically related to cloze value as opposed to occurring randomly, so allowing false hearing to vary randomly across items may suppress the effect of cloze value. This analysis was conducted once using data from the SPIN task in Experiment 1 and again using data from the SPIN task in Experiment 2. For the analysis of Experiment 1 data, a dummy-coded counterbalancing variable and the interaction of counterbalancing condition with cloze value were also included as predictors to account for differences in average cloze value across counterbalancing conditions.

Looking first at the data from the SPIN task in Experiment 1, I found that the interaction between cloze value and age group was non-significant, so I dropped the interaction term from the model to allow for interpretation of main effects. In the updated model, the interaction between cloze value and counterbalancing condition was significant (*OR* = 1.31, $z = 3.85$, $p < .001$), so I analyzed the effect of age group and cloze value separately within each counterbalancing condition. The interaction of age group and cloze value was not significant in the first counterbalancing condition (*OR* = .95, $z = -.47$, $p > .05$) or the second counterbalancing condition (*OR* = .85, $z = -1.61$, $p > .05$), so the interaction terms were dropped from the models. In the first counterbalancing condition, a one standard deviation increase in cloze value increased the odds of false hearing by a factor of 1.13, which was significant ($z = 2.37$, $p < .05$). Age group was also a marginally significant predictor of false hearing after controlling for differences in cloze value in the first counterbalancing condition (*OR* = 1.46, $z = 1.79$, $p < .07$), suggesting that older adults tended to be more susceptible to false hearing than younger adults controlling for the cloze value of sentences. In the second counterbalancing condition, cloze value had an even greater effect on susceptibility to false hearing: A one standard deviation increase in cloze value increased the odds of false hearing by a factor of 1.53 ($z = 7.78$, $p < .001$). Age group was not a significant predictor of false hearing after controlling for differences in cloze value in the second counterbalancing condition (*OR* = 1.31, $z = 1.00$, $p > .05$). Thus, although the magnitude of the effect of cloze value differed across counterbalancing conditions[12], more predictive sentences were more likely to evoke false hearing in Experiment 1.

Interestingly, the interaction between cloze value and age group was marginally significant in Experiment 2 (*OR* = .77, z = -1.99, p = .05). To investigate this interaction, I tested the relationship between cloze value and false hearing separately in each age group. Younger adults in Experiment 2 looked much like the participants in Experiment 1, with a one standard deviation increase in cloze value increasing the odds of false hearing by a factor of 1.39 (z = 4.07, $p < .001$). However, there was no significant relationship between cloze value and false hearing for older adults (*OR* = 1.09, z = .81, $p > .05$). Thus, the predictive strength of contextual cues was positively related to susceptibility to false hearing in all cases except for older adults in Experiment 2. A possible reason for the absence of a relationship between cloze value and susceptibility to false hearing for older adults in Experiment 2 is detailed in the Experiment 2 discussion.

The use of the hear/know/guess paradigm in the SPIN task from Experiment 1 also allowed me to test whether sentences that were more predictive were more likely to result in false hearing experiences characterized as heard relative to known or guessed. To test this hypothesis, a mixed-effects multinomial logistic regression analysis was conducted using the brm function from the *brms* package in R (Bürkner, 2017). Age group, standardized cloze value, the age group by standardized cloze value interaction, counterbalancing condition, and the counterbalancing condition by standardized cloze value interaction were entered as predictors.

---

[12] The magnified effect of cloze value in the second counterbalancing condition may have occurred because sentences were slightly more predictive on average in this condition relative to the first counterbalancing condition (.78 vs .76).

The odds of characterizing a response as heard relative to either known or guessed and the effect of cloze value on those odds were allowed to vary randomly across subjects.

In the mixed-effects multinomial logistic regression model, the age group by standardized cloze value interaction was non-significant when comparing both know and hear judgments (*OR* = 1.07, *95% CI* = .80 – 1.41) and guess and hear judgements (*OR* = 1.10, *95% CI* = .69 – 1.73), so it was removed as a predictor. The interaction between counterbalancing condition and cloze value was also non-significant when comparing guess and hear judgements (*OR* = 1.09, *95% CI* = .69 – 1.75) but was significant when comparing know and hear judgements (*OR* = .68, *95% CI* = .51 – .89), so this interaction was left in the model. The counterbalancing condition by cloze value interaction remained significant when comparing know and hear judgements in the updated model (*OR* = .68, *95% CI* = .51 – .90), so I conducted the analysis again within each counterbalancing condition.

In the first counterbalancing condition, the odds of characterizing a false hearing response as known relative to heard increased significantly as cloze value increased (*OR* = 1.26, *95% CI* = 1.04 – 1.56), an effect in the opposite direction of what I had predicted. Cloze value was unrelated to the odds of characterizing a false hearing response as guessed relative to heard (*OR* = .95, *95% CI* = .57 – 1.79). In this counterbalancing condition, older adults were less likely than younger adults to characterize false hearing responses as known relative to heard controlling for cloze value (*OR* = .50, *95% CI* = .29 – .83), but there was no difference between age groups for the odds of characterizing false hearing responses as guessed relative to heard (*OR* = 1.61, *95% CI* = .50 – 5.28).

In the second counterbalancing condition, the odds of characterizing a false hearing response as known relative to heard did not vary significantly with cloze value (*OR* = .85, *95%*

CI = .66 – 1.08), and the same was true for characterizing a false hearing response as guessed relative to heard (*OR* = .93, *95% CI* = .64 – 1.44). Older adults were again less likely than younger adults to characterize false hearing responses as known relative to heard controlling for cloze value (*OR* = .39, *95% CI* = .17 – .92), and there was no difference between age groups for the odds of characterizing false hearing responses as guessed relative to heard (*OR* = .63, *95% CI* = .26 – 1.58). Therefore, across counterbalancing conditions, I did not observe the predicted increase in false hearing responses characterized as heard relative to known or guessed as sentences became increasingly predictive, and I observed the opposite effect for heard versus known responses in the first counterbalancing condition. A potential reason for these unpredicted effects is discussed below.

3.2.5 Fixation analyses

To determine changes in the proportion of fixations on each image across time in the visual word task, linear mixed-effects regression was used to analyze the proportion of fixations on each image following the analyses used in a recent eye-tracking study that employed a similar visual world paradigm (Ito et al., 2018). Separate analyses were conducted for each sentence type (baseline, congruent, incongruent). Sentences were divided into three 2000 ms bins for fixation analyses. The first bin started from 500 ms after the start of the trial since there were very few fixations on any of the images before this time (participants tended to still be looking at the central fixation cross). The second time bin started 2500 ms into the trial and continued until just before the target word was presented. The third time bin started 4500 ms into the trial, exactly when the target word onset. These three bins allowed me to determine the proportion of fixations on each image early in the sentence (bin 1), late in the sentence but before the target word was presented (bin 2), and from the presentation of the target word onwards (bin 3). Each mixed-effects model had 24 dummy coded variables corresponding to the four picture types (target image, alternative image, and two foil images) within each age group (younger/older adults) at each time bin as fixed effects predicting the proportion of fixations. Proportion of fixations was allowed to vary randomly across subjects and items for each image type within each time bin. Linear combinations of the fixed effects were tested using the *multcomp* package in R (Hothorn et al., 2008) to determine how the proportion of fixations on each image changed over time, and whether these changes differed across age groups. Since the target image and the alternative image (i.e., the contextually predicted image in incongruent sentences) were of primary interest to my hypotheses, I will focus on fixation trends for these image types in the Results section. For analysis of the fixation trends for the two unrelated foil images, see the Appendix.

3.2.5.1 Baseline sentences

The proportion of fixations over time for baseline sentences is presented in Figure 21. For baseline sentences, I predicted that fixations on the target image would not increase until time bin 3 since there were no contextual cues in baseline sentences to afford anticipatory activation to any particular response. This prediction was supported by the fixation analysis. In baseline sentences, there was no difference in fixations on the target image from time bin 1 to time bin 2 ($ED = .03$, $z = 1.48$, $p > .05$), but fixations on the target image increased from time bin 2 to time bin 3 ($ED = .27$, $z = 12.80$, $p < .001$). There was no interaction with age group for the difference from time bin 1 to time bin 2 ($ED = .00$, $z = .08$, $p > .05$) or from time bin 2 to time bin 3 ($ED = .02$, $z = .74$, $p > .05$). Therefore, as predicted, both younger and older adults only increased fixations on the target image after the target word had been presented in baseline sentences, demonstrating that neither group experienced anticipatory activation of the target word when no contextual cues were present.

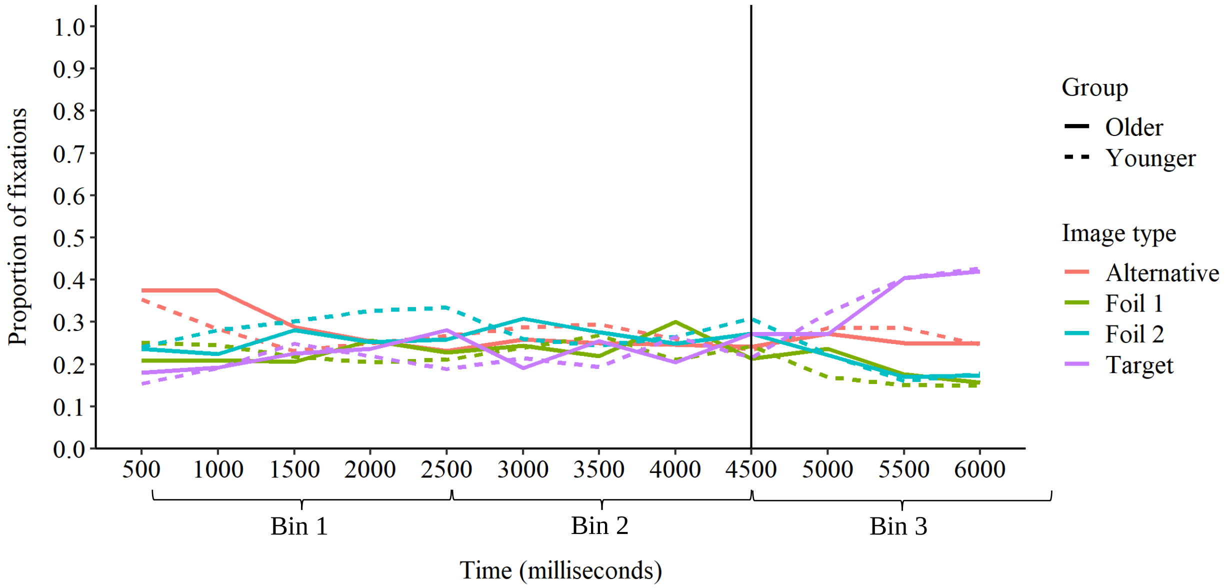

Figure 21. Proportion of fixations on each image type over time in baseline sentences for younger and older adults. The vertical line at 4500 ms represents the onset of the target word.

I next considered changes in the proportion of fixations on the alternative image (i.e., the image depicting the phonological neighbor of the target word). In baseline sentences, there was a significant reduction in fixations on the alternative image from time bin 1 to time bin 2 (*ED* = -.05, *z* = -2.24, *p* < .05). However, there was an interaction with age group (*ED* = .05, *z* = 2.68, *p* < .01) that revealed that older adults reduced fixations on the alternative image from time bin 1 to time bin 2 (*ED* = -.05, *z* = -3.29, *p* < .01), whereas younger adults did not (*ED* = .00, *z* = .32, *p* > .05) . From time bin 2 to time bin 3, there was a marginally significant decrease in proportion of fixations on the alternative image (*ED* = -.04, *z* = -1.76, *p* = .08) and there was no significant interaction with group (*ED* = -.03, *z* = -1.22, *p* > .05), demonstrating that the proportion of fixations on alternative image decreased slightly as fixations on the target image continued to increase for both groups.

3.2.5.2 Congruent sentences

For congruent sentences, I had predicted that both age groups would begin looking towards the contextually predicted target image before the target word was presented. Additionally, I predicted that older adults might increase their fixations on the target image to a greater degree than younger adults, reflecting increased influence of context over responding. As can be seen in Figure 22, these predictions were mostly supported by the fixation data. Fixations on the target image increased from time bin 1 to time bin 2 ($ED = .25$, $z = 12.98$, $p < .001$) and again from time bin 2 to time bin 3 ($ED = .26$, $z = 12.81$, $p < .001$). Whereas there was no interaction between age group and the change in fixations from time bin 1 to time bin 2 ($ED = -.01$, $z = -.27$, $p > .05$), there was an interaction with age group for the change in fixations from time bin 2 to time bin 3 ($ED = -.15$, $z = -7.17$, $p < .001$). For younger adults, there was a significant increase in proportion of fixations on the target image from time bin 2 to time bin 3 ($ED = .06$, $z = 4.13$, $p < .001$), but this increase was much greater for older adults ($ED = .20$, $z = 13.69$, $p < .001$). Thus, both younger and older adults increased fixations on the target image before the target word was presented, demonstrating anticipatory activation of the target word based on available contextual cues. Older adults increased fixations on the target image to a greater degree than younger adults, but only after the target word had been presented. This suggests that younger and older adults formed context-based expectations at a similar rate, but older adults became more fixated on the contextually predicted response once additional support for this response was provided by presentation of the target word. The greater increase in fixations on the target image after the target word was presented by older, relative to younger, adults suggests that older adults may use the auditory signal to confirm their context-based expectations. When the auditory signal supports the word they expected to hear, older adults become increasingly fixated on that response option, whereas younger adults may be more cautious and consider alternative options.

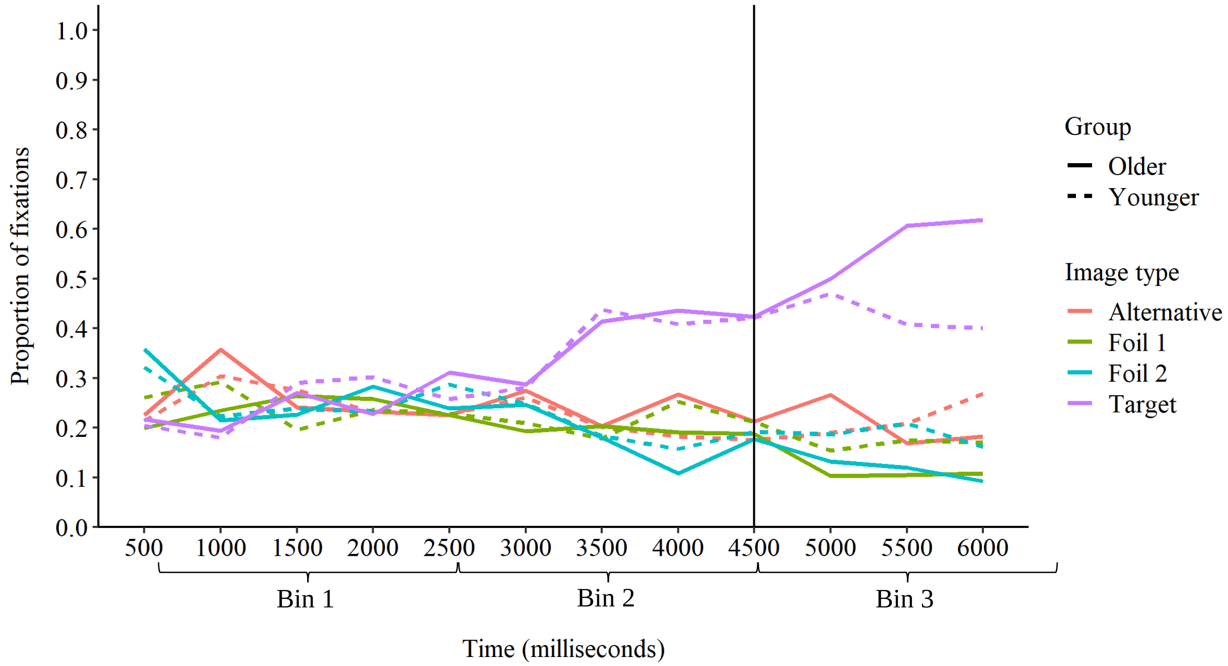

Figure 22. Proportion of fixations on each image type over time in congruent sentences for younger and older adults. The vertical line at 4500 ms represents the onset of the target word.

As fixations increased on the target image, the proportion of fixations on the alternative image decreased. Fixations decreased from time bin 1 to time bin 2 for the alternative image (*ED* = -.08, *z* = -4.13, *p* < .001) and there was no interaction with group (*ED* = -.02, *z* = -.84, *p* > .05). Fixations did not decrease significantly from time bin 2 to time bin 3 for the alternative image (*ED* = -.02, *z* = -.96, *p* > .05). However, there was a significant interaction with age group (*ED* = .06, *z* = 3.02, *p* < .01) indicating that younger adults did not reduce their fixations on the alternative image from time bin 2 to time bin 3 (*ED* = .02, *z* = 1.51, *p* > .05), whereas older adults did (*ED* = -.04, *z* = -2.73, *p* < .01).

3.2.5.3 Incongruent sentences

Finally, for incongruent sentences, I had again predicted that both younger and older adults would look towards the contextually predicted image before the target word was presented. In incongruent sentences, the alternative word was predicted by context as opposed to the target word, so I predicted that fixations on the alternative image would increase before the target word was presented. Here again I had predicted that older adults would increase fixations on the alternative image to a greater degree than younger adults, reflecting greater influence of context over responding. After the target word was presented, I predicted that fixations on the alternative image would decrease and fixations on the target image would increase for younger adults as they realized that the presented target word differed from the contextually predicted word. However, I predicted that older adults would be less likely than younger adults to shift their focus towards the target image, instead either maintaining or increasing fixations on the alternative image, reflecting older adults' poorer ability to suppress the expected response and increased susceptibility to false hearing in incongruent sentences.

Average fixations over time in incongruent sentences can be seen in Figure 23. As predicted, the proportion of fixations on the alternative image increased from time bin 1 to time bin 2 ($ED = .17$, $z = 8.87$, $p < .001$), whereas fixations on the target image did not increase ($ED = -.02$, $z = -1.07$, $p > .05$). There was no interaction with age group for the difference between time bin 1 and time bin 2 for fixations on the alternative image ($ED = -.01$, $z = -.70$, $p > .05$), but there was a marginally significant interaction for fixations on the target image ($ED = -.04$, $z = -1.92$, $p = .05$). This interaction reflected that younger adults significantly decreased fixations on the target image from time bin 1 to time bin 2 ($ED = -.03$, $z = -2.24$, $p < .05$) whereas older adults' fixations on the target image did not change ($ED = .01$, $z = .57$, $p > .05$). Interestingly, fixations increased from time bin 2 to time bin 3 for both the alternative image ($ED = .04$, $z = 2.03$, $p < .05$) and the target image ($ED = .16$, $z = 7.65$, $p < .001$). However, as predicted, there were significant interactions with age group for the difference in fixations from time bin 2 to time bin 3 for both the alternative image ($ED = -.14$, $z = -6.90$, $p < .001$) and the target image ($ED = .12$, z

= 5.88, *p* < .001). Younger adults decreased fixations on the alternative image (*ED* = -.05, *z* = -3.51, *p* < .001) and increased fixations on the target image (*ED* = .14, *z* = 9.75, *p* < .001) from time bin 2 to time bin 3. Older adults, however, increased fixations on the alternative image (*ED* = .09, *z* = 6.20, *p* < .001) and did not increase fixations on the target image (*ED* = .02, *z* = 1.23, *p* > .05) from time bin 2 to time bin 3. As can be seen in Figure 23, older adults only began to slightly decrease fixations on the alternative image and increase fixations on the target image at the very end of time bin 3, whereas younger adults made this change in fixations shortly after the target word had been presented.

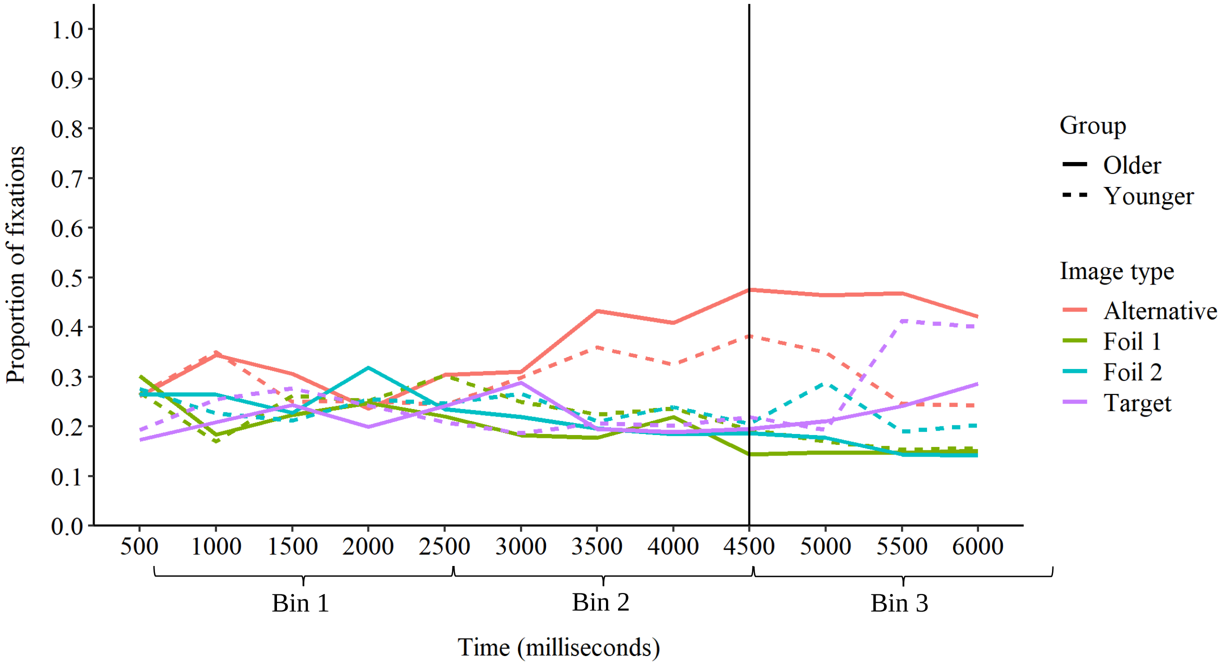

Figure 23. Proportion of fixations on each image type over time in incongruent sentences for younger and older adults. The vertical line at 4500 ms represents the onset of the target word.

3.3 Experiment 2 discussion

The findings of Experiment 2 were interesting for several reasons. This was the first study of false hearing to employ a four-alternative forced-choice test as opposed to using open-set responding. By having participants click on one of four images to indicate which word they heard in noise as opposed to saying the word aloud as in previous studies (Rogers et al., 2012; Sommers et al., 2015) or typing the word as in Experiment 1, I limited the range of words that participants had to consider in the perceptual process. This resulted in a substantially easier task, requiring the SNRs to be decreased to get similar baseline accuracy to that in Experiment 1. However, despite having an image depicting the target word on screen in incongruent sentences, participants – especially older adults – still experienced false hearing and even dramatic false hearing. This demonstrates the robust effects of context on speech perception: Even when the right answer was presented to participants in an image, younger adults still chose the contextually predicted (but incorrect) option on approximately 50% of incongruent trials, whereas older adults committed these context-based errors on approximately 70% of incongruent trials.

This was also the first study to test the relationship between the strength of contextual cues and susceptibility to false hearing, a relationship that has been observed in past studies of false memory (Deese, 1959; Roediger et al., 2001b). As in false memory, the predictive strength of sentences in the SPIN task was positively related to the likelihood of experiencing false hearing in both Experiment 1 and Experiment 2 of the present study. This finding supports an activation-based account of false hearing. Specifically, this finding lends credence to Sommers and Danielson's (1999) suggestion that hearing predictive contextual cues increases the activation of words in the mental lexicon that are compatible with the context. However, this finding expands upon Sommers and Danielson's hypothesis by suggesting that the activation allotted to contextually predicted words is proportional to the predictive strength of the sentence. Thus, misleading sentences with greater predictive strength allotted more activation to the contextually predicted (but unpresented) word, resulting in increased probability that the activation of the contextually predicted word would exceed that of the target word, resulting in false hearing.

Interestingly, the relationship between cloze value and susceptibility to false hearing was not observed for older adults in Experiment 2. One possible reason for this null finding is that only one image was compatible with context on each incongruent trial. Even in cases where the context provided by incongruent sentences was a relatively weak predictor, the alternative image (e.g., the image depicting a box in the sentence *She put the toys in the* <u>*fox*</u>) was still the only one that fit with the semantic structure of the sentence. So, if older adults were more likely to respond based on context than younger adults, we would expect that they would select the contextually compatible image even in cases where context was a relatively weak predictor. By contrast, the open-set response format in Experiment 1 would allow older adults to generate semantically viable words other than the pre-selected alternative word, and thus weaker contexts would correspond to a lower probability of generating the specific alternative word. This would account for the positive relationship between cloze value and susceptibility to false hearing for older adults observed in Experiment 1 and the absence of a significant relationship in Experiment 2.

Although greater predictive strength of sentences was related to increased susceptibility to false hearing, it was not related to increased odds of characterizing false hearing responses as heard relative to known or guessed in Experiment 1. In fact, I found that the odds of characterizing false hearing responses as known relative to heard increased as predictive strength increased in one counterbalancing condition. It is possible that these unpredicted findings resulted from restriction of range of cloze values. Given that the above findings established that sentences with greater predictive strength were more likely to evoke false hearing than those with weaker predictive strength, it is possible that including only data in which false hearing occurred limited the number of sentences with weaker predictive strength in the analysis. To test this possibility, I conducted the same mixed-effects multinomial regression model including all incongruent trials as opposed to just cases of false hearing to determine whether responses for incongruent sentences with stronger predictive strength were more likely to be rated as heard relative to known or guessed. It is important to note that this analysis included correct responses on incongruent trials, which would be predicted to occur more often for sentences with low cloze values (since sentences with high cloze values were more likely to result in false hearing) and to be rated as heard (since the auditory signal was the only source of information supporting this response). Thus, including these trials actually worked against my hypothesis that increasing cloze value should increase the proportion of responses rated as heard on incongruent trials.

As in the original analysis, the age group by cloze value interaction was non-significant for the comparison of incongruent responses rated as heard versus known (*OR* = .95, *95% CI* = .80 – 1.13) and heard versus guessed (*OR* = 1.07, *95% CI* = .91 – 1.27), so it was removed from the model to allow for interpretation of the main effects. The interaction between cloze value and counterbalance condition was significant for the comparison of responses rated as heard versus guessed (*OR* = 1.31, *95% CI* = 1.11 – 1.56), so I tested the effect of age group and cloze value within each counterbalancing condition. In the first counterbalancing condition, cloze value was unrelated to the odds of characterizing responses as known versus heard (*OR* = 1.05, *95% CI* = .90 – 1.22), but more predictive sentences were less likely to be characterized as guessed than heard (*OR* = .63, *95% CI* = .54 – .72). There were no age effects for the odds of classifying responses as known relative to heard (*OR* = 1.18, *95% CI* = .66 – 2.08) or guessed relative to heard (*OR* = .86, *95% CI* = .49 – 1.47). In the second counterbalancing condition, cloze value was unrelated to the odds of characterizing responses as known versus heard (*OR* = .94, *95% CI* = .83 – 1.06), but more predictive sentences were again less likely to be characterized as guessed than heard (*OR* = .82, *95% CI* = .73 – .93). There were no age effects for the odds of classifying responses as known relative to heard (*OR* = 1.04, *95% CI* = .53 – 1.98) or guessed relative to heard (*OR* = 1.03, *95% CI* = .65 – 1.60) in the second counterbalancing condition. Therefore, as misleading sentences became more predictive and participants became more likely to commit a context-based error, the odds that participants would rate their responses as heard as opposed to guessed increased. This suggests that increasing the predictive strength of sentences altered the subjective experience associated with responses to be more similar to veridical hearing.

Finally, I presented eye-tracking data that provided the first visualization of the real-time processing that occurred during baseline, congruent, and incongruent sentences. I found that participants formed early expectations regarding what the target word would be in the congruent and incongruent conditions, increasing their fixations on the image supported by context before the target word was presented. Additionally, whereas older adults' gaze tended to linger on the contextually predicted (but incorrect) image after the target word was presented in incongruent sentences, younger adults were more likely to shift their gaze to the correct target image. These findings align with the idea that one can be “captured” by available contextual cues, and that older adults are especially prone to being captured. These findings also support the role of the response suppression aspect of inhibitory control in false hearing, suggesting that older adults were less able than younger adults to suppress the highly activated (but incorrect) response in the incongruent condition.

One prediction that I had made at the outset of the study that was not reflected in the eye-tracking data was increased expectation on the part of older adults relative to younger adults. Specifically, I predicted that older adults would increase fixations on the contextually predicted image to a greater degree than younger adults before the target word was presented in congruent and incongruent sentences. I found that older adults did display a greater proportion of fixations on the contextually predicted image, but only after the target word had been presented. This was true in both the congruent and incongruent conditions, leading to increased fixations on the correct image when context was valid and increased fixations on the incorrect image when context was misleading for older, relative to younger, adults. This finding suggests that younger and older adults formed expectations based on context at a similar rate, but older adults became disproportionately fixated on the contextually predicted image after the target word was presented. As described above, this may suggest that older adults used the auditory signal to confirm their context-based expectations, whereas younger adults were more willing to consider alternative responses to that predicted by context.

Chapter 4: General discussion

In the present study, I sought to learn more about the mechanisms underlying false hearing by drawing upon the more expansive false memory literature and using a diverse set of research methods, including behavioral assessments of performance, self-reports of subjective experience, analyses of individual differences, MPT modeling, and eye-tracking. It has been suggested that false hearing and false memory may rely upon similar cognitive mechanisms (Jacoby et al., 2012; Rogers et al., 2012; Sommers et al., 2015), but this was the first study to directly test this theory. I tested whether individual differences in two cognitive processes previously shown to be predictive of false memory – inhibitory control (Colombel et al., 2016; Sommers & Huff, 2003, Experiment 2) and executive functioning (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007) – were also predictive of false hearing. I then further examined the role of inhibitory control by testing whether the predictive strength of sentences was related to susceptibility to false hearing. Finally, I used eye-tracking to determine whether participants became increasingly fixated on the contextually predicted response over the course of sentences, and whether younger and older adults differed in their ability to suppress the activation of the predicted response when context turned out to be invalid.

4.1 Relationship between false hearing and false memory

Supporting the idea that false hearing and false memory share at least one underlying cognitive mechanism, I found a significant positive relationship between susceptibility to false hearing and susceptibility to false memory. Importantly, I also found that age group was not a significant predictor of false hearing after controlling for an individual's average susceptibility to false memory. This finding supports the theory that age differences in a common cognitive mechanism, as opposed to age differences in sensory-specific or memory-specific mechanisms, gave rise to older adults' increased susceptibility to false hearing and false memory relative to younger adults.

Whereas past studies have hypothesized that false hearing and false memory might rely on common cognitive mechanisms (Jacoby et al., 2012; Rogers et al., 2012; Sommers et al., 2015), the present study provided the first direct evidence for a relationship between the two phenomena. Showing that false hearing and false memory are correlated does not, however, identify what common mechanism(s) gave rise to these phenomena. To begin to investigate specific mechanisms, I tested the effects of two related cognitive abilities that have been mentioned in past studies of false hearing as potential mechanisms underlying false hearing and false memory: inhibitory control and executive functioning.

4.2 Evidence for mechanisms underlying the relationship between false hearing and false memory

4.2.1 Inhibitory control

To test the role of inhibitory control in false hearing and false memory, I first employed an auditory Stroop task that has previously been shown to be predictive of phonological false memory in the DRM paradigm (Sommers & Huff, 2003, Experiment 2). I did not replicate the relationship between the response suppression aspect of inhibitory control and false memory observed by Sommers and Huff, nor did I observe a relationship between this aspect of inhibitory control and false hearing. This suggests that inhibitory control as assessed by the auditory Stroop task was not the common mechanism underlying false hearing and false memory. However, as stated above, it is possible that this null finding resulted from insensitivity of the auditory Stroop task to age differences in inhibitory control, lack of validity of the auditory Stroop task as a measure of inhibitory control, different demands placed on inhibitory control in the auditory Stroop task relative to the SPIN and memory tasks, or my use of a different memory task than Sommers and Huff. It is also plausible that the auditory Stroop task measured a different aspect of inhibitory control (see Lustig et al., 2007) than was involved in the incongruent conditions SPIN and memory tasks. For example, as will be described in further detail below, the results of the cloze and eye-tracking analyses suggest that false hearing results when participants are unable to suppress a highly activated (but incorrect) response, but interference on the auditory Stroop task may reflect the increased effort needed to stop the task-irrelevant spoken word from entering working memory. Future studies should investigate which aspect of inhibitory control is measured by the auditory Stroop task to clarify this issue.

To further assess the role of inhibitory control in false hearing in the present study, I analyzed the relationship between the predictive strength of sentences (i.e., cloze values) and susceptibility to false hearing. Previous studies using the DRM paradigm have shown that the associative strength of studied list words to the critical word is highly predictive of false memory (Deese, 1959; Roediger et al., 2001b). Roediger et al. (2001b) interpreted their findings within an activation/monitoring framework, which proposes that false memories arise when 1) an incorrect response becomes highly activated in memory due to its association with contextual cues, and 2) the individual is unable to distinguish whether this activation reflected a veridical memory trace or resulted from context (i.e., a failure of source monitoring). The authors argued that the stronger the association between the studied list words and the critical word, the more activation should spread from the list items to the critical word, and that the likelihood of falsely remembering the critical word should increase as the word became more highly activated. Within this framework, inhibitory control can be used to suppress the spread of activation from studied items to the unpresented critical word (Balota et al., 1999; Roediger et al., 2001a), and thus, individual differences in inhibitory control may be related to individual differences in susceptibility to false memory.

This activation-based explanation for false memory is similar to an explanation I described above for false hearing based on the expanded version of the NAM (Luce & Pisoni, 1998) proposed by Sommers and Danielson (1999, Experiment 2). Specifically, Sommers and Danielson suggested that inhibitory control is used to suppress activation of phonological neighbors of the spoken word in the mental lexicon, and that placing the spoken word within a predictive semantic context reduces the need to inhibit semantically incongruent neighbors by selectively increasing the activation of the spoken word. I argued that false hearing may occur when a phonological neighbor of the spoken word is predicted by context whereas the spoken word is not – as in the incongruent condition of the SPIN task – thereby selectively increasing the activation of the contextually predicted phonological neighbor. This would increase reliance on inhibitory control to suppress activation of contextually predicted phonological neighbors to accurately hear the spoken word. In cases where the listener was unable to suppress the activation of the contextually predicted neighbor to a level less than the activation of the spoken word, false hearing could occur. Therefore, just as increasing activation of the critical word by strengthening its association with studied words increased susceptibility to false memory (Deese, 1959; Roediger et al., 2001b), increasing the activation of a phonological neighbor of a spoken word by strengthening contextual cues predicting this neighbor may also increase susceptibility to false hearing. This was the first study to test this hypothesis.

Whereas analyses using the auditory Stroop task as a predictor of false hearing and false memory suggested that the response suppression aspect of inhibitory control was not related to either phenomenon, the relationship between this aspect of inhibitory control and false hearing was supported by the cloze value analyses. Sentences that were stronger predictors of a phonological neighbor of the spoken target word were more likely to evoke false hearing than were less predictive sentences. This supports the hypothesis that false hearing occurs when context increases the activation of a phonological neighbor of the target word beyond the activation of the target word. However, given that even the most predictive sentences (*Use a comb to fix your…* and *He got a letter in the…* each with cloze values of 1.00 indicating that all participants completed the sentences with the same words, *hair* and *mail*) did not evoke false hearing in all cases (Hair: $M_{Younger}$ = .27 and M_{Older} = .45; Mail: $M_{Younger}$ = .14 and M_{Older} = .21), participants may have been able to inhibit the activation gained from context to accurately hear the target word. Therefore, the cloze value analyses suggest that failure to sufficiently inhibit the activation of contextually predicted phonological neighbors of the spoken word in the mental lexicon may contribute to false hearing. This explanation aligns with activation-based theories of false memory (Balota et al., 1999; Roediger et al., 2001a).

The role of the response suppression aspect of inhibitory control in false hearing and the activation-based explanation of false hearing presented above are supported by the fixation data from the visual world paradigm used in Experiment 2. Increasing fixations on a particular image may represent increasing activation of the word that the image depicts (Tanenhaus et al., 2000). In incongruent sentences, the proportion of fixations on the contextually predicted alternative image increased as the sentence was played for both younger and older adults, suggesting that activation of the alternative image increased above the level of the other three response options because it alone fit with the semantic context. Once the target word was presented in the incongruent condition, younger adults decreased their fixations on the alternative image and increased fixations on the target image, representing their ability to inhibit activation of the contextually predicted alternative word to accurately hear the target word. Older adults, on the other hand, did not decrease their fixations on the alternative image or increase their fixations on the target image to the same extent as did younger adults, suggesting that older adults were less able to inhibit the activation of the contextually predicted alternative word than younger adults. This is in line with past findings of poorer inhibitory control in older, relative to younger, adults (Cohn et al., 1984; Hasher & Zacks, 1988; Hasher et al., 1997; Jacoby et al., 2005; MacLeod, 1991; Sommers & Danielson, 1999, Experiment 2; Sommers & Huff, 2003, Experiment 2; West

& Alain, 2000), and would account for older adults' increased susceptibility to false hearing relative to younger adults. Thus, whereas the auditory Stroop task was unable to detect age differences in inhibitory control in the present study, it is possible that these differences were captured by fixation trends in the visual world task.

As a final assessment of the role of inhibitory control in false hearing, I fit the capture model used by Jacoby et al. (2005, Experiment 2) to the SPIN and memory task data in Experiment 1. I found that adding a capture parameter – interpreted by Jacoby et al. as representing inhibitory control – to models significantly improved overall fit to data from both the memory and SPIN tasks. This finding was important, as it suggested that the model that accounted for hearing/recollecting the target word, accessibility-based responding, and guessing was insufficient to accurately predict response rates. An additional capture parameter, allowing for increased rates of both congruent hits and cases of false hearing and false memory in the incongruent condition, improved the model's predictive ability. Although it is possible that the capture parameter could reflect processes other than, or in addition to, inhibitory control that would result in similar patterns of responding, finding that adding the capture parameter improved model fit, in conjunction with evidence from the cloze and eye-tracking analyses, suggests that inhibitory control may indeed have played a role in both the memory and SPIN tasks.

One interesting finding from the MPT models was that the best-fitting model to the SPIN data described in the Experiment 1 discussion did not specify age differences in the capture parameter. This seems to contradict the conclusion drawn from the fixation analyses and suggests that younger and older adults did not differ in inhibitory control. There are several potential explanations for these findings. First, it is possible that changing the paradigm from open-set responding in Experiment 1 to four-alternative forced-choice in Experiment 2 increased the need to employ inhibitory control. It could be the case that having an image of the predicted (but incorrect) word on screen on incongruent trials in Experiment 2 increased participants' focus on that response option, thereby requiring increased inhibitory control to suppress this response. This explanation is supported by substantially higher rates of false hearing for both younger and older adults in Experiment 2 (see Figure 19) relative to Experiment 1 (see Figure 8). Second, it is possible that the magnitude of age differences in the capture parameter was suppressed by the capture parameter's inclusion of both correct responses in the congruent condition and false hearing responses in the incongruent condition. Since age differences appeared to be smaller for congruent hits relative to cases of false hearing (see Figure 8), potentially due to near-perfect performance by both groups in the congruent condition, it is possible that including congruent hits reduced the ability of the model to detect age differences in the capture parameter. Indeed, allowing the capture parameter to vary across age groups in this model revealed that older adults were numerically more susceptible to capture than younger adults (.32 vs. .26), but the BIC value indicated that this difference was not large enough to justify the added model complexity. Therefore, both the fixation and MPT analyses are suggestive of age differences in inhibitory control, but these differences were not large enough to justify the added model complexity in the MPT analysis.

Although the MPT model provided additional evidence for the role of inhibitory control in false hearing, it is important to note that the capture model I originally tested based on the model used by Jacoby et al. (2005, Experiment 2) tended to overestimate the number of cases of false hearing rated as heard or known. It was not until individual accessibility bias parameters were specified in each sentence condition that the model was able to accurately estimate cases of false hearing. This suggests that accessibility bias – the tendency to respond with the most easily retrieved response – may play a role in false hearing. Additionally, whereas the best-fitting model for the memory task specified age group differences in the capture parameter, the best-fitting model for the SPIN task did not specify group differences in capture but did specify group differences in accessibility bias. This suggests that age differences in susceptibility to false hearing may reflect, in part, older adults' increased bias towards the most easily accessible word relative to younger adults.

Importantly, age differences in accessibility bias can also account for the results of the cloze value analyses and the age differences in fixation patterns observed in the visual world task in Experiment 2. Above, I argued that strengthening predictability of sentences increased susceptibility to false hearing because the increased activation afforded by context made it more difficult to inhibit the contextually predicted (but unpresented) word. However, it could also be the case that more predictive sentences were more likely to result in false hearing because these sentences made the predicted word more easily accessible than other options. Participants may thus have chosen this response in cases where they were unable to hear the target word because it was the most easily accessed option, not because they were unable to resist this response. Similarly, I explained above that older adults' inability to reduce fixations on the contextually predicted image after the target word was presented in the incongruent condition in Experiment 2 reflected a failure to inhibit the expected response. However, it could also be the case that older adults did not decrease fixations on the expected image whereas younger adults did because older adults were more heavily biased towards the most easily accessible option, whereas younger adults were more likely to consider the less easily accessible target word. Therefore, both inhibitory control and accessibility bias can explain these findings.

Age differences in accessibility bias may also help to explain why older adults increased fixations on the target image to a greater degree than younger adults after the target word was presented in the congruent condition. Above, I argued that this increase may result because older adults used the auditory signal to confirm their context-based expectation, becoming increasingly fixated on the target response when it was supported by both context and the auditory signal. I suggested that this increase in fixations on the target image may have been smaller in younger adults because younger adults were more willing to consider alternative options. These patterns can also be explained in terms of accessibility bias. According to the expanded version of the NAM (Luce & Pisoni, 1998) proposed by Sommers and Danielson (1999), activation of words in the mental lexicon is increased by both phonological similarity to the spoken word and congruency with available context. Therefore, we would expect activation of the target word to increase as contextual support is presented in congruent sentences, and that its activation would also increase when the target word is presented due to phonological similarity, further increasing the accessibility of this response. Older adults' increased bias towards the most easily accessible option would lead them to become more fixated on the target image after the target word was presented, whereas younger adults may consider alternative options because they are less biased towards the most easily accessible option. Therefore, differences in fixation patterns across younger and older adults in both the congruent and incongruent conditions can be readily explained by differences in accessibility bias.

The similarity of the effects of inhibitory control and accessibility bias is reflected in the capture model itself. As can be seen in Figure 4, responding based on capture and accessibility bias both result in accurate responding in the congruent condition and false hearing in the incongruent condition. Additionally, like capture, responses based on accessibility bias can be characterized as heard/remembered or known in cases where the attribution threshold is met. Thus, it is unsurprising that the inhibitory control and accessibility bias accounts are difficult to disentangle.

The idea that differences in accessibility bias may contribute to age differences in false hearing is not a new one. In discussing the high rate of accurate responses in the congruent condition and high rate of false hearing responses in the incongruent condition of their study, Rogers et al. (2012) stated that the “increase in both hits and false alarms provides evidence that the effect of providing context was, at least, partially due to an influence on bias” (pp. 42). The authors went on to suggest that age differences in false hearing resulted from decreased reliance on controlled, effortful processing (i.e., careful listening) and increased reliance on automatic, context-based responding by older, relative to younger, adults. Although Sommers et al. (2015) described older adults’ increased reliance on automatic processing as potentially resulting from failures of inhibitory control (i.e., failing to restrict responding to sensory information), it is possible that this reflects a shift towards increased accessibility-based responding in older, relative to younger, adults: Rather than expending the extra effort to listen carefully to the stimuli in noise, older adults may simply provide the response made easily accessible by context. Similarly, it is possible that the increased reliance on automatic memory processes and context-based guessing attributed to older adults with low relative to high executive functioning in past studies (Chan & McDermott, 2007; Meade et al., 2012) reflects a change in accessibility bias. Thus, differences in accessibility bias could account for both age differences in susceptibility to false hearing (Rogers et al., 2012; Sommers et al., 2015) and differences in susceptibility to false memory attributed to executive functioning in past studies (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007).

As indicated within the capture model itself, both inhibitory control and accessibility bias can independently contribute to false hearing. Indeed, demonstrating that model fit was improved by adding a capture parameter to an MPT model that already included an accessibility bias parameter suggests that capture – which may represent failure to inhibit the response activated by context – accounted for unique variability in responding in both tasks. Even after specifying a separate hearing parameter in the baseline condition and a separate accessibility bias parameter in the congruent condition in the model described in the Experiment 1 discussion, model fit was significantly better when the capture parameter was included relative to when it was constrained to 0 ($\Delta G^2(1) = 167.32$, critical = 36.08). Similarly, it is possible that both inhibitory control and accessibility bias influenced the effect of sentence predictability (i.e., cloze value) on false hearing, and that age differences in each of these processes contributed to older adults' inability relative to younger adults to reduce fixations on the predicted image in the incongruent condition of the visual word task. Future studies should seek to dissociate the influences of inhibitory control and accessibility bias on false hearing and false memory.

4.2.2 Executive functioning

Whereas findings with regards to inhibitory control were mixed, I found that better executive functioning was associated with decreased susceptibility to both false hearing and false memory. Importantly, the relationship between executive functioning and false hearing remained significant after controlling for best-ear PTA, suggesting that false hearing does not purely reflect deficits in hearing acuity. Additionally, I found that age group was not a significant predictor of either false hearing or false memory after controlling for performance on the executive functioning battery, suggesting that differences in executive functioning can fully account for age differences in each of these phenomena. This corroborated the findings of past studies showing that age differences in susceptibility to false memory are reduced or eliminated when comparing younger adults to only older adults with high executive functioning (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007), and suggested that a common role of executive functioning may account for the observed relationship between false memory and false hearing.

At the outset of the study, I had predicted that executive functioning would not account for additional variance in susceptibility to false hearing and false memory after controlling for individual differences in the response suppression aspect of inhibitory control. This prediction was based on the fact that executive functioning is composed of several cognitive processes, including inhibitory control, and the common theory that false hearing results from failures to inhibit the context-based response (Rogers et al., 2012; Sommers et al., 2015). Therefore, it is possible that the relationship between executive functioning and false memory observed in past studies (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007) may, in fact, be indexing a relationship between the response suppression aspect of inhibitory control and false memory.

To attempt to assess whether executive functions other than the response suppression aspect of inhibitory control were related to susceptibility to false hearing and false memory, I conducted an analysis of the relationship between the executive functioning battery and each of these phenomena controlling for performance on the auditory Stroop task. Unfortunately, due to concerns surrounding the auditory Stroop task as a measure of the response suppression aspect of inhibitory control described in detail above, I was not able to draw conclusions regarding the roles of other executive functions in false hearing and false memory in the present study. However, showing that executive functioning, in general, was related to both false hearing and false memory is important to advancing the theory that these phenomena rely on similar cognitive mechanisms, and narrows the scope of potential mechanisms to be investigated in greater detail in future studies.

Although I focused on inhibitory control as a specific component of executive functioning that may explain the relationship between executive functioning and false memory in past studies (Butler et al., 2004; Chan & McDermott, 2007; Meade et al., 2012; Roediger & Geraci, 2007) and between executive functioning and both false hearing and false memory here, it is possible that other aspects of executive functioning contribute to these phenomena. For example, past studies have argued that different response strategies may explain differences in susceptibility to false memory between individuals with low and high executive functioning (Chan & McDermott, 2007; Meade et al., 2012). Above, I argued that this difference in strategies could reflect either differences in inhibitory control or accessibility bias. Additionally, there is some evidence that working memory capacity may play a role in false memory (Bixter & Daniel, 2013; Long et al., 2008; Unsworth & Brewer, 2010; Watson et al., 2005). Specifically, past studies have shown that individuals with greater working memory capacity were less likely to endorse lures as “old” on recognition tests (Long et al., 2008; Unsworth & Brewer, 2010) and less likely to experience false memories in the DRM paradigm (Bixter & Daniel, 2013; Watson et al., 2005) than individuals with lower working memory capacity. However, individuals with greater working memory capacity were only less susceptible to false memories in the DRM paradigm when participants were warned that the word lists were designed to elicit false memories (Bixter & Daniel, 2013; Watson et al., 2005). This led the authors of these studies to suggest that working memory capacity is involved in maintaining task goals as opposed to having a direct effect on susceptibility to false memory. It is important that future studies investigate the contributions of individual components of executive functioning to false hearing and false memory to better understand why these phenomena occur and to develop methods of reducing susceptibility to these errors.

4.3 Implications for the common-mechanisms theory of false hearing and false memory

The findings presented in this study provide substantial evidence in support of the theory that false hearing and false memory result from at least one shared cognitive mechanism. The findings that most strongly supported this theory were the demonstration of a positive correlation between susceptibility to false hearing and false memory, and the finding of a positive relationship between executive functioning and both false hearing and false memory. I not only presented the first direct evidence of a relationship between false hearing and false memory, suggesting that a common mechanism may give rise to these phenomena, but also presented evidence that executive functioning may be one such mechanism (or set of mechanisms).

Further evidence for the common-mechanisms theory came from the demonstration that false hearing was more likely to occur as the predictive strength of sentences increased, similar to the finding of a positive relationship between backwards associative strength and false memory observed in previous studies (Deese, 1959; Roediger et al., 2001b). Showing that similar manipulations – in this case, increasing predictive strength of contextual cues – has a similar effect on false hearing and false memory strengthens the case that similar processes underlie these phenomena. I argued that either inhibitory control, accessibility bias, or a combination of the two could account for the effect of increasing predictive strength on susceptibility to false hearing, and the same could be argued for its effect on susceptibility to false memory. Above, I described how inhibitory control may be needed to suppress spreading activation from studied list words to the critical word in the DRM paradigm, and how failures to suppress this activation may result in false memory. However, it could also be argued that participants endorse the critical word in the DRM paradigm because it has been made easily accessible in memory due to its association with the studied list words, not because they are unable to resist this response. Therefore, the similar effect of increasing predictive strength of contextual cues on false hearing and false memory could be explained by similar roles of inhibitory control, accessibility bias, or a combination of both.

This study also provided evidence for differences between false hearing and false memory. For example, initial attempts to fit the MPT model revealed that this model was a better fit to the memory task data than to the SPIN data: Additional parameters (a different hearing parameter in the baseline condition and a different accessibility bias parameter in the congruent condition) had to be specified to achieve accurate estimations of the observed SPIN data. This suggests that the effect of context may have differed in the memory and SPIN tasks, resulting in different patterns of responding that could not be accounted for with the same model. The MPT models also suggested that age differences in responding in the memory and SPIN tasks were best accounted for by different parameters: capture and recollection in the memory task, and accessibility bias and word generation in the SPIN task. This suggests that whereas similar processes may have contributed to responding in each task, different processes may have given rise to age differences in responding in the memory and SPIN tasks.

There was also evidence that the subjective experience associated with false memories differed from that associated with cases of false hearing. Specifically, whereas the vast majority of cases of false hearing were characterized as heard, there was relatively equal likelihood of characterizing false memories as remembered, known, or guessed. This could suggest that the subjective experience of false hearing was more similar to that of veridical hearing than the subjective experience of false memory was to that of veridical memory. However, as described above, it is also possible that characteristics of the specific memory task used in the present study (e.g., the ease of generating the false memory response provided by the remaining letters in the fragment completion task) gave rise to the high rate of false memories characterized as guesses, rather than reflecting a characteristic common to all false memories.

Finally, showing that the relationship between age and false hearing persisted after controlling for average susceptibility to false memory suggests that age differences in mechanisms that are not shared with false memory contributed to age differences in false hearing. For example, it could be the case that sensory-specific mechanisms, such as hearing acuity, contributed to susceptibility to false hearing but did not impact susceptibility to false memory. Conversely, memory-specific mechanisms, such as the ability to encode information into long-term memory and the ability to retrieve this information, may have contributed to susceptibility to false memory without affecting susceptibility to false hearing. The findings of the present study provide compelling evidence for at least one shared mechanism, but also suggest that additional, distinct mechanisms contribute to false hearing and false memory.

4.4 Implications for our understanding of speech perception

In addition to informing the theory that false memory and false hearing share at least one common mechanism, the findings of the present study also advance our understanding of speech perception. The results of the cloze value analyses, for example, have important implications for activation-based theories of speech perception. Whereas Sommers and Danielson (1999) suggested that activation of words in the mental lexicon is greater when placed in the context of a highly predictive sentence relative to a sentence providing no predictive value, the results of the cloze value analyses extended this argument by suggesting that the activation imparted by context varies continuously with the predictive strength of contextual cues. This finding will allow for better prediction of the outcome of perception by models such as the NAM (Luce & Pisoni, 1998).

The fixation analyses from Experiment 2 also offered interesting insights into how younger and older adults process contextual cues in speech perception. Specifically, these analyses showed that both younger and older adults formed expectations regarding what would be said in the future based on the semantic cues presented earlier in a sentence. Younger and older adults formed these expectations at a similar rate, which countered my expectation that older adults might become fixated on the expected word to a greater degree than younger adults, reflecting increased reliance on context as a basis for responding. It was only after the target word had been presented that older adults demonstrated increased fixation on the expected word relative to younger adults: Older adults increased fixations on the expected word to a greater degree than younger adults after presentation of the target word in congruent sentences and failed to reduce fixations on the expected word in the incongruent condition to the same degree as younger adults. This pattern of fixations after presentation of the target word suggests that older adults may have used the target word to confirm, but not to override, their context-based expectation. When the target word aligned with their expectation (i.e., in the congruent condition), older adults increased their fixations on the expected image. When the target word differed from their expectation (i.e., in the incongruent condition), older adults did not decrease their fixations on the expected image. This lends further credence to the idea that context plays a greater role in determining the outcome of perception for older, relative to younger, adults

(Rogers et al., 2012; Sommers et al., 2015), and as argued above, may reflect increased bias towards the most easily accessible response by older adults.

Finally, this study was the first to assess the subjective experience of cases of false hearing, giving us greater insight into whether participants actually “hear” the incorrect word or are simply making an informed guess based on context. To assess the subjective experience of false hearing, I used a hear/know/guess paradigm – similar to the remember/familiar/guess paradigm used in the study of false memory by Jacoby et al. (2005) – wherein participants indicated whether they had heard the specific sounds of the word when it was presented, had not heard the specific sounds of the word but knew their answer was the word that had been presented, or had simply guessed. The findings of this study suggest that cases of false hearing are accompanied by a subjective experience similar to that of veridical hearing, with the vast majority of cases of false hearing being classified as heard. This not only advances our understanding of false hearing, but also advances our understanding of speech perception, generally, suggesting that the subjective experience of hearing is not always based on one’s ability to discern the acoustic features of a spoken word, but can also be based on one’s expectations regarding what word *should* have been presented.

4.5 Limitations and future directions

There are several limitations associated with the present study. One limitation is the use of a single task (the auditory Stroop task) as opposed to a composite of multiple tasks to measure the response suppression aspect of inhibitory control in Experiment 1. Above and beyond the potential issues with the auditory Stroop task discussed above, using a composite of multiple tasks assessing a particular construct provides a better measure than a single task since it allows the researcher to partial out variance that is not shared across measures. This removes error variance and task-specific variance, thereby providing a more reliable measure of the latent construct. For example, Colombel et al. (2016) used a composite of three measures of inhibitory control – a Stroop task (Stroop, 1935), the Hayling task (Burgess & Shallice, 1997), and a directed forgetting task (see Sego et al., 2006) – to test the relationship between inhibitory control and false memory in the DRM paradigm.

There were two primary reasons for my decision to use only the auditory Stroop task as a measure of the response suppression aspect of inhibitory control in the present study. First, one of the primary goals of this study was to determine whether measures and manipulations that have been shown to be related to susceptibility to false memory in past studies were also related to susceptibility to false hearing. To accomplish this goal, I used the exact measures that have previously demonstrated relationships with false memory. Since performance on the auditory Stroop task was shown to be related to susceptibility to false memory by Sommers and Huff (2003, Experiment 2), I wanted to determine whether this measure was also predictive of false hearing. The second reason I decided to use only the auditory Stroop task was that I was concerned that participants would experience cognitive fatigue if more tasks were added to the procedure. Participants completed 10 distinct tasks split across two one-hour sessions in Experiment 1. I was worried that adding more tasks would cause participants – older adults in particular – to lose focus, resulting in data that was not representative of their cognitive ability and defeating the purpose of having multiple measures of inhibitory control. However, in future studies that focus specifically on inhibitory control, I recommend that researchers use multiple measures of inhibitory control to increase the reliability of their measure.

A second limitation of Experiment 1 is that I only focused on false memories evoked using the process dissociation paradigm (see Jacoby, 1991; Jacoby et al., 2005). Many other paradigms have been used to study false memory, including the DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), misinformation paradigms (Loftus, 1975), inference paradigms (Brewer, 1977; Chan & McDermott, 2006; Chan & McDermott, 2007), schema paradigms (Bartlett, 1932; Brewer & Treyens, 1981), and the imagination inflation paradigm (Goff & Roediger, 1998). It is currently unclear whether false memories resulting from these diverse paradigms result from similar mechanisms. As mentioned above, there is some evidence showing that susceptibility to false memories evoked using the DRM paradigm is uncorrelated with susceptibility to false memories evoked using a misinformation paradigm (Ost et al., 2013; Patihis et al., 2018) and that there is no significant correlation between false memories evoked using DRM paradigms with semantically related lists and those with phonologically related lists (Ballou & Sommers, 2008), suggesting that different mechanisms may be involved in different types of false memories even within the same basic paradigm. Therefore, whereas I have shown that a correlation exists between susceptibility to false hearing and susceptibility to false memory evoked using a process dissociation paradigm, it is not necessarily the case that susceptibility to false hearing is related to susceptibility to false memories evoked using different paradigms. Similarly, as mentioned above, it is possible that my inability to replicate Sommers and Huff's (2003, Experiment 2) finding of a relationship between the response suppression aspect of inhibitory control and false memory resulted from my use of a process dissociation paradigm and their use of a phonological DRM paradigm. Therefore, researchers should be cautious when extrapolating the findings of this study to other types of false memory.

The final limitation that I will consider here is the truncated sample size in Experiment 2. Experiment 2 was originally intended to have a sample of 30 younger and 30 older adults, but data collection was halted due to the COVID-19 pandemic. The resulting sample size was approximately two-thirds of what was intended. Because of this limited sample size, it is possible that results could change in a larger sample. However, given that the data followed all predictions made at the outset of the study – including replicating the patterns in the SPIN task from Experiment 1 – and that observed differences were statistically reliable, I believe that the conclusions drawn are valid. Nonetheless, I recommend that future studies seek to replicate the findings from Experiment 2 with a larger sample.

One reason, in particular, that it would be useful to replicate Experiment 2 with a larger sample is to allow for more detailed analysis of fixations over time in the visual world paradigm. My limited sample resulted in random fluctuations in the proportion of fixations, especially early in sentences, that may have been smoothed in a larger sample size. To counteract this issue, I decided to group fixations into three relatively large (2000 ms) time bins to simplify analyses. While binning data in this way still allowed me to address my hypotheses, future studies with more fixation data could potentially examine changes in the proportion of fixations on each image in bins of 500 ms or shorter. This added level of detail could improve our understanding of age differences in context-use in speech perception and of sentence processing, generally.

4.6 Conclusions

The present study offered several insights into false hearing and speech processing in general. This was the first demonstration of the often-hypothesized relationship between false hearing and false memory, furthering the theory that these phenomena rely on similar cognitive mechanisms. Importantly, this study also offered evidence for a specific mechanism – or rather, a set of mechanisms – that could underlie this relationship: executive functioning. Showing that individuals with low executive functioning were more prone to both false hearing and false memory advances our theoretical understanding of these phenomena and helps to identify individuals who may be most at risk of experiencing them.

In addition to identifying the individuals who were most susceptible to false hearing, Experiment 2 offered insights into the stimuli that were most likely to give rise to false hearing. I found that the greater the predictive strength of a sentence, the greater the likelihood that false hearing would occur. Using eye-tracking, I was able to show how participants – older adults in particular – became increasingly focused on the contextually supported response as predictive sentences were presented. Older adults were less likely than younger adults to shift their gaze from the contextually predicted word to the target word when context was misleading. These findings were interpreted as evidence for the potential roles of the response suppression aspect of inhibitory control and accessibility bias in speech perception. Predictive sentence context may have increased activation of the expected word in the mental lexicon, resulting in an incorrect response becoming highly activated when context was misleading. False hearing then occurred when participants were either unable to suppress the activation of the contextually predicted word or were biased towards responding with the most easily accessible response. Regardless of the mechanism (or combination of mechanisms), these findings suggest that it is important to be mindful of preceding information and the expectations of the listener when conversing, as it may be difficult for listeners to abandon these expectations in cases where they are violated.

Finally, it is important to note that context is rarely misleading outside of the laboratory, and thus, context-based responding is adaptive. Rather than seeing older adults' increased susceptibility to false hearing and false memory as a cognitive deficit, it could be viewed as the result of an effective adaptation to changing life circumstances. The perceptual system may recalibrate to emphasize easily accessible, context-based responses to compensate for declining hearing acuity across the lifespan. The results of this study increase our understanding of this adaptation and should be used to improve communication with older adults. Knowing that older adults – particularly those with low executive functioning – are prone to overuse of contextual cues, speakers could increase their likelihood of being understood by carefully framing important information within supporting context. Providing this type of scaffolding may improve both initial perception and later memory of the presented information.

References

1. Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439. doi:10.1006/jmla.1997.2558

2. Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 20,* 385-406. doi:10.1207/S15326942DN2001_5

3. Baddeley, A. D., & Hitch, G. (1974). Working memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 47-87). Academic. doi:10.1016/S0079-7421(08)60452-1

4. Ballou, M. R., & Sommers, M. S. (2008). Similar phenomena, different mechanisms: Semantic and phonological false memories are produced by independent mechanisms. *Memory & Cognition, 36,* 1450-1459. doi:10.3758/MC.36.8.1450

5. Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., III, McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology, 16,* 361-384. doi:10.1080/026432999380834

6. Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445-459. doi:10.3758/BF03193014

7. Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

8. Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1-48. doi:10.18637/jss.v067.i01

9. Benichov, J., Cox, L. C., Tun, P. A., & Wingfield, A. (2012). Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability and cognitive function. *Ear Hear, 32,* 250-256. doi:10.1097/AUD.0b013e31822f680f

10. Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32-48. doi:10.1121/1.2017541

11. Bixter, M. T., & Daniel, F. (2013). Working memory differences in illusory recollection of critical lures. *Memory & Cognition, 41,* 716-725. doi:10.3758/s13421-013-0293-x

12. Boone, K. B., Pontón, M. O., Gorsuch, R. L., González, J. J., & Miller, B. L. (1998). Factor analysis of four measures of prefrontal lobe functioning. *Archives of Clinical Neuropsychology, 13,* 585-595. doi:10.1016/S0887-6177(97)00074-7

13. Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5*, 673-678. doi:10.3758/BF03197414

14. Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 207-230. doi:10.1016/0010-0285(81)90008-6

15. Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Thames Valley Test Company.

16. Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1-28. doi:10.18637/jss.v080.i01

17. Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L., III (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review, 11*, 921-925. doi:10.3758/BF03196722

18. Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology, 20*, 633-639. doi:10.1002/acp.1215

19. Chan, J. C. K., & McDermott, K. B. (2007). The effects of frontal lobe functioning and age on veridical and false recall. *Psychonomic Bulletin & Review, 14*, 606-611. doi:10.3758/BF03196809

20. Cohn, N. B., Dustman, R. E., & Bradford, D. C. (1984). Age-related decrements in Stroop Color Test performance. *Journal of Clinical Psychology, 40*, 1244-1250. doi:10.1002/1097-4679(198409)40:5<1244::AID-JCLP2270400521>3.0.CO;2-D

21. Colombel, F., Tessoulin, M., Gilet, A.-L., & Corson, Y. (2016). False memories and normal aging: Links between inhibitory capacities and monitoring processes. *Psychology and Aging, 31*, 239-248. doi:10.1037/pag0000086

22. Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language, 57*, 483-501. doi:10.1016/j.jml.2007.01.001

23. Dahan, D., Magnuson, J. S., Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, 317-367. doi:10.1006/cogp.2001.0750

24. Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22. doi:10.1037/h0046671

25. Dobie, R. A. (2011). The AMA method of estimation of hearing disability: A validation study. *Ear and Hearing, 32*, 732-740. doi:10.1097/aud.0b013e31822228be

26. Dubno, J. R., Ahlstrom, J. B., Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *Journal of the Acoustical Society of America, 107*, 538-546. doi:10.1121/1.428322

27. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. doi:10.3758/BRM.41.4.1149

28. Finley, J. R., Sungkhasettee, V. W., Roediger, H. L., III, & Balota, D. A. (2017). Relative contributions of semantic and phonological associates to over-additive false recall in hybrid DRM lists. *Journal of Memory and Language, 93*, 154-168. doi:10.1016/j.jml.2016.07.006

29. Garner, W. R. (1974). *The processing of information and structure*. Erlbaum.

30. Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9*, 229-235. doi:10.1037/0894-4105.9.2.229

31. Goff, L. M., & Roediger, H. L., III (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26*, 20-33. doi:10.3758/BF03211367

32. Hart, R. P, Kwentus, J. A., Wade, J. B., & Taylor, J. R. (1988). Modified Wisconsin Sorting Test in elderly normal, depressed and demented patients. *The Clinical Neuropsychologist, 1*, 49-56. doi:10.1080/13854048808520085

33. Hasher, L., Quig, M. B., & May, C. P. (1997). Inhibitory control over no-longer relevant information: Adult age differences. Memory & Cognition, 25, 286-295. doi:10.3758/BF03211284

34. Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193-225. doi:10.1016/S0079-7421(08)60041-9

35. Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185. doi:10.1007/BF02289447

36. Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346-363. doi:10.1002/bimj.200810425

37. Huff, M. J., & Umanath, S. (2018). Evaluating susceptibility to additive and contradictory misinformation following explicit error detection in younger and older adults. *Journal of Experimental Psychology: Applied, 24*, 180-195. doi:10.1037/xap0000138

38. Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journal of Gerontology: Psychological Sciences, 44*, 36-44. doi:10.1093/geronj/44.2.p36

39. Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1-11. doi:10.1016/j.jml.2017.09.002

40. Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541. doi:10.1016/0749-596X(91)90025-F

41. Jacoby, L. L. (1999). Deceiving the elderly: Effects of accessibility bias in cued-recall performance. *Cognitive Neuropsychology, 16*, 417-436. doi:10.1080/026432999380861

42. Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General, 134*, 131-148. doi:10.1037/0096-3445.134.2.131

43. Jacoby, L. L., Rogers, C. S., Bishara, A. J., & Shimizu, Y. (2012). Mistaking the recent past for the present: False seeing by older adults. *Psychology and Aging, 27*, 22-32. doi:10.1037/a0025924

44. Jacoby, L. L., Wahlheim, C. N., Rhodes, M. G., Daniels, K. A., & Rogers, C. S. (2010). Learning to diminish the effects of proactive interference: Reduce false memory for young and older adults. *Memory & Cognition, 38*, 820-829. doi:10.3758/MC.38.6.820

45. Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functioning: A review of our current understanding. *Neuropsychological Review, 17*, 213-233. doi:10.1007/s11065-007-9040-z

46. Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151. doi:10.1177/001316446002000116

47. Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology, 16*, 399-415. doi:10.1080/026432999380852

48. Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9*, 491-512. doi:10.1037/0882-7974.9.4.491

49. Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi:10.1037/xlm0000935

50. Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*, 560-572. doi:10.1016/0010-0285(75)90023-7

51. Long, D. L., Prat, C., Johns, C., Morris, P., & Jonathan, E. (2008). The importance of knowledge in vivid text memory: An individual-differences investigation of recollection and familiarity. *Psychonomic Bulletin & Review, 15*, 604-609. doi:10.3758/PBR.15.3.604

52. Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*, 268-283. doi:10.1016/S0749-596X(03)00069-X

53. Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear Hear, 19*, 1-36. doi:10.1097/00003446-199802000-00001

54. Lustig, C., Hasher, L., & Zacks, R. (2007). Inhibitory deficit theory: Recent developments in a “new view.” In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 145-162). American Psychological Association.

55. MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*, 163-203. doi:10.1037/0033-2909.109.2.163

56. Mazaheri, M., Hoogkamer, W., Potocanac, Z., Verschueren, S., Roerdink, M., Beek, P. J., Peper, C. E., & Duysens, J. (2015). Effects of aging and dual tasking on step adjustments to perturbations in visually cued walking. *Experimental Brain Research, 233*, 3467-3474. doi:10.1007/s00221-015-4407-5

57. McCabe, D. P., Roediger, H. L., III, McDaniel, M. A., & Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering: Neuropsychological test correlates of remember-know judgments. *Neuropsychologia, 47*, 2164-2173. doi:10.1016/j.neuropsychologia.2008.11.025

58. McCabe, D. P., Roediger, H. L., III, McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*, 222-243. doi:10.1037/a0017619

59. McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition, 34*, 1273-1284. doi:10.3758/bf03193271

60. Meade, M. L., Geraci, L. D., & Roediger, H. L., III (2012). Neuropsychological status in older adults influences susceptibility to false memories. *American Journal of Psychology, 125*, 449-467. doi:10.5406/amerjpsyc.125.4.0449

61. Millar, P. R., Balota, D. A., Bishara, A. J., & Jacoby, L. L. (2018). Multinomial models reveal deficits of two distinct controlled retrieval processes in aging and very mild Alzheimer disease. *Memory & Cognition, 46*, 1058-1075. doi:10.3758/s13421-018-0821-9

62. Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *International Journal for Phonetic Science, 41*, 215-225. doi:10.1159/000261728

63. Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word process in Hindi-English sequential bilinguals: an eye tracking visual world study. *Reading and Writing: An Interdisciplinary Journal, 27*, 129-151. doi:10.1007/s11145-013-9436-5

64. Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology, 17*, 107-119. doi:10.1002/acp.857

65. Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., & Fozard, J. L. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *Journal of the Acoustical Society of America, 100,* 1949-1967. doi:10.1121/1.417906

66. Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America, 87*, 2705-2715. doi:10.1121/1.399061

67. Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25,* 838-848. doi:10.3758/BF03211328

68. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences, 16*, 292-305. doi:10.1016/j.tics.2012.04.005

69. Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). False memory ≠ false memory: DRM errors are unrelated to the misinformation effect. *PLoS ONE, 8*: e57939. doi:10.1371/journal.pone.0057939

70. Patihis, L., Frenda, S. J., & Loftus, E. F. (2018). False memory tasks do not reliably predict other false memories. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 140-160. doi:10.1037/cns0000147

71. Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*, 593-608. doi:10.1121/1.412282

72. Pineda, D. A., & Merchan, V. (2003). Executive function in young Colombian adults. *International Journal of Neuroscience, 113*, 397-410. doi:10.1080/00207450390162164

73. Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55-85). Lawrence Erlbaum.

74. Prull, M. W., & Yockelson, M. B. (2013). Adult age-related differences in the misinformation effect for context-consistent and context-inconsistent objects. *Applied Cognitive Psychology*, *27,* 384-395. doi:10.1002/acp.2916

75. Raiche, G., & Magis, D. (2020). *Parallel analysis and other non graphical solutions to the Cattell Scree Test (Version 2.4.1)*. https://cran.r-project.org/web/packages/nFactors/nFactors.pdf

76. R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/

77. Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging, 27,* 80-87. doi:10.1037/a0024113

78. Rodríguez-Aranda, C., & Sundet, K. (2006). The frontal hypothesis of cognitive aging: Factor structure and age effects on four frontal tests among healthy individuals. *The Journal of Genetic Psychology, 167,* 269-287. doi:10.3200/GNTP.167.3.269-287

79. Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001a). Spreading activation and arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). American Psychological Association. doi:10.1037/10394-006

80. Roediger, H. L., III, & Geraci, L. (2007). Aging and the misinformation effect: A neuropsychological analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 321-334. doi:10.1037/0278-7393.33.2.321

81. Roediger, H. L., III, & McDaniel, M. A. (2007). Illusory recollection in older adults: Testing Mark Twain's conjecture. In M. Gerry & H. Hayne (Eds.), *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom* (pp. 105-136). Erlbaum.

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

83. Roediger, H. L., III, Watson, J. M., McDermott. K. B., & Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8,* 385-407. doi:10.3758/BF03196177

84. Rogers, C. S., Jacoby, L. L., Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging, 27,* 33-45. doi:10.1037/a0026231

85. Sego, S. A., Golding, J. M., & Gottlob, L. R. (2006). Directed forgetting in older adults using the item and list methods. *Neuropsychology, Development, and Cognition: A Journal on Normal and Dysfunctional Development, 13,* 95-114. doi:10.1080/138255890968682

86. Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9,* 371-377. doi:10.1080/00223980.1940.9917704

87. Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of Multinomial Processing Tree models in R. *Behavior Research Methods, 45,* 560-575. doi:10.3758/s13428-012-0259-0

88. 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,* 458-472. doi:10.1037/0882-7974.14.3.458

89. Sommers, M. S., Hale, S., Myerson, J., Rose, N., Tye-Murray, N., & Spehar, B. (2011). Listening comprehension across the adult lifespan. *Ear and Hearing, 32*, 775-781. doi:10.1097/aud.0b013e3182234cf6

90. Sommers, M. S., & Huff, L. M. (2003). The effects of age and dementia of the Alzheimer's type on phonological false memories. *Psychology and Aging, 18*, 791-806. doi:10.1037/0882-7974.18.4.791

91. Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40*, 83-108. doi:10.1006/jmla.1998.2614

92. Sommers, M. S., Morton, J., & Rogers, C. (2015). You are not listening to what I said: False hearing in young and older adults. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III (Eds.), *Remembering: Attributions, Processes, and Control in Human Memory (essays in Honor of Larry Jacoby)* (pp. 269-284). Psychology Press.

93. Spreen, O., & Benton, A. L. (1977). *Neurosensory Center comprehensive examination for aphasia* (rev. ed.). University of Victoria, Neuropsychology Laboratory.

94. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. doi:10.1037/h0054651

95. Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557-580. doi:10.1023/A:1026464108329

96. Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1-12. doi:10.1037/h0080017

97. Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62*, 19-34. doi:10.1016/j.jml.2009.08.002

98. Vervoort, D., den Otter, A. R., Buurke, T. J. W., Vuillerme, N., Hortobágyi, T., & Lamoth, C. J. C. (2019). Effects of aging and task prioritization on split-belt gait adaptation. *Frontiers in Aging Neuroscience, 11*, 1-12. doi:10.3389/fnagi.2019.00010

99. Watson, J. M., Balota, D. A., & Roediger, H. L., III (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language, 49*, 95-118. doi:10.1016/S0749-596X(03)00019-6

100. Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology, 15*, 254-267. doi:10.1037//0894-4105.15.2.254

101. Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 76-85. doi:10.1037/0278-7393.31.1.76

102. Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised manual*. Psychological Corp.

103. Wechsler, D. (1997). *Wechsler Memory Scale – III manual*. Psychological Corp.

104. West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology, 37*, 179-189. doi:10.1111/1469-8986.3720179

105. Wingfield, A., Aberdeen, J. S., & Stine, E. A. (1991). Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology, 46,* P127-P129. doi:10.1093/geronj/46.3.P127

106. Wingfield, A., Tun, P. A., Koh, C. K., & Rosen, M. J. (1999). Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging, 14,* 380-389. doi:10.1037/0882-7974.14.3.380

107. Zwick, W. R., & Velicer W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442. doi:10.1037/0033-2909.99.3.432

Appendix

Here, I present analyses of the proportion of fixations on the two foil images – which I will refer to as foil 1 and foil 2 – over the course of baseline, congruent, and incongruent sentences.

Baseline

Fixations did not change from time bin 1 to time bin 2 for either foil 1 ($ED = .00$, $z = .18$, $p > .05$) or foil 2 ($ED = .01$, $z = .60$, $p > .05$). From time bin 2 to time bin 3, fixations decreased significantly for both foil 1 ($ED = -.12$, $z = -5.69$, $p < .001$) and foil 2 ($ED = -.11$, $z = -5.35$, $p < .001$) as fixations on the target image increased. There were no significant interactions with group (ps > .05).

Congruent

Fixations decreased from time bin 1 to time bin 2 for foil 1 ($ED = -.05$, $z = -2.39$, $p < .05$) and foil 2 ($ED = -.13$, $z = -6.46$, $p < .001$). There was no interaction with group for foil 1 ($ED = -.02$, $z = -.79$, $p > .05$), but there was a marginally significant interaction for foil 2 ($ED = .04$, $z = 1.90$, $p = .06$). Younger adults decreased fixations on foil 2 from time bin 1 to time bin 2 ($ED = -.04$, $z = -3.41$, $p < .001$) but older adults decreased their fixations on this image to a greater degree ($ED = -.08$, $z = -5.61$, $p < .001$). Fixations decreased further from time bin 2 to time bin 3 for foil 1 ($ED = -.13$, $z = -6.41$, $p < .001$) and foil 2 ($ED = -.11$, $z = -5.44$, $p < .001$). There were significant interactions with age group for the change in fixations between time bin 2 and time bin 3 for foil 1 ($ED = .06$, $z = 3.04$, $p < .01$) but not for foil 2 ($ED = .02$, $z = -1.11$, $p > .05$). For foil 1, younger adults reduced their fixations from time bin 2 to time bin 3 ($ED = -.03$, $z = -2.47$, $p < .05$) but older adults reduced their fixations to an even greater degree ($ED = -.10$, $z = -6.48$, $p < .001$).

Incongruent

As fixations increased on the alternative image and, later, the target image, there was a steady decrease in fixations on the two foil images in incongruent sentences. For foil 1, fixations decreased from time bin 1 to time bin 2 (*ED* = -.07, *z* = -3.53, *p* < .001) and from time bin 2 to time bin 3 (*ED* = -.11, *z* = -5.40, *p* < .001). Fixations on foil 2 also decreased from time bin 1 to time bin 2 (*ED* = -.08, *z* = -4.28, *p* < .001) and from time bin 2 to time bin 3 (*ED* = -.09, *z* = -4.28, *p* < .001). There were no significant interactions with age group (*p*s > .05).