THE EFFECTS OF OUTPUT INTERFERENCE ON METAMEMORY AND CUED RECALL ACCURACY IN YOUNG AND OLDER ADULTS

A Dissertation
Presented to
The Academic Faculty

By

Taylor M. Curley

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology
College of Sciences

Georgia Institute of Technology

December 2021

THE EFFECTS OF OUTPUT INTERFERENCE ON METAMEMORY AND CUED RECALL ACCURACY IN YOUNG AND OLDER ADULTS

Thesis committee:

Dr. Christopher Hertzog, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Dobromir Rahnev
School of Psychology
Georgia Institute of Technology

Dr. Rick Thomas
School of Psychology
Georgia Institute of Technology

Dr. John Dunlosky
Department of Psychological Sciences
Kent State University

Dr. Paul Verhaeghen
School of Psychology
Georgia Institute of Technology

Date approved: August 25, 2021

ACKNOWLEDGMENTS

Thank you to my friends and family, and especially Mom, Dad, Caitlin, Alex, and Philip, who provided love and support throughout my dream to earn my doctorate—I love you all very much. Thank you also to my committee—Rick Thomas, Paul Verhaeghen, Doby Rahnev, and John Dunlosky—for all of your help and encouragement throughout the years, and especially during this thesis. I also want to thank everybody in the Adult Cognition Lab. I'm especially grateful to the undergraduates who helped me throughout the years—Jayna, Omer, Hannah, Kirsten, Josh, Kenley, Alysha, Skyler, Aliyah, Aiman, Mackenzie, Faizah, Yusra, Ana, and Caroline—and were absolutely essential to my success in graduate school. Importantly, thank you to my lab mates—Emily, MacKenzie, Maugan, and Marit—who provided me with all of the encouragement, help, and friendship that I could ever ask for. Finally, I want to thank my advisor, Chris Hertzog, for his encouragement, excellent guidance, and sage advice throughout my years at Georgia Tech and particularly during this dissertation. I truly owe my success to these wonderful people.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

γ or *G* Goodman-Kruskal Gamma Correlation Coefficient

4AFC 4-alternative Forced Choice Test

dJOL Delayed Judgment of Learning

FOK Post-recall Feeling-of-knowing Judgment

HyGene Hypothesis Generation Model

iJOL Immediate Judgment of Learning

LTM Long-term Memory

OI Output Interference

PI Proactive Interference

POK Predictions of Knowing

R/K/N Remember/Know/No Memory Judgments

RCJ Post-recognition Confidence Judgment

RI Retroactive Interference

RIF Retrieval-induced Forgetting

WM/WMC Working Memory/Working Memory Capacity

SUMMARY

Output interference (OI) is a gradual decline in memory accuracy as a function of an item's position in a testing sequence (M. C. Anderson & Neely, 1996). Despite having been researched for over 50 years (e.g., Tulving & Arbuckle, 1963), this effect has yet to be linked to metacognitive experiences. The current study examines differences in memory accuracy and monitoring for young and older adults who experience OI during cued recall. At study, participants were asked to remember 40 cue-target pairs: For half of the participants, cue words were exemplars that were sampled from the same taxonomic category, while for the other half of the participants, word pairs were completely unrelated. At test, participants first engaged in a cued recall task, where they were asked to predict future recognition memory outcomes (i.e. feelings-of-knowing; FOKs) as well as if they experienced feelings of "Remembering", "Knowing", or "No Memory" (i.e. R/K/N judgments) for each trial. Afterwards, participants engaged in a 4-alternative forced-choice recognition task and were asked to provide retrospective confidence judgments (RCJs) after each trial. In the aggregate, memory and metamemory accuracy were similar for young and older adults in both experimental conditions. At the level of the trial, however, recall accuracy, FOKs, and self-reported recollection significantly decreased across successive trials for participants of all ages experiencing OI. Decreases in memory accuracy during OI were mirrored by increases in retrieval failures and states of no memory. Only self-reported familiarity differed between age groups, where "Know" judgments decreased across trials for young adults, but increased for older adults. The results support previous findings of age invariance in FOK accuracy (Hertzog, Sinclair, et al., 2010) and highlight the role of retrieval suppression in mechanistic accounts of OI.

CHAPTER 1

INTRODUCTION

Adults often show increased difficulties in remembering learned information across the lifespan (Craik, 1994; Light, 1991; Schaie, 2005). This is often compounded by negative self-beliefs older adults have about their own memory abilities and control over such abilities (Dixon & Hultsch, 1983; Hertzog, Small, McFall, & Dixon, 2019; Lineweaver & Hertzog, 1998); however, older adults' beliefs about their memory abilities may not match true memory performance (Hertzog et al., 2018).

A goal of metamemory research in adulthood is determining the extent to which memory judgment accuracy changes over the lifespan—if at all. In particular, the degree to which adults differ with respect to *feelings-of-knowing*, or predictions that one will be able to recognize an item that she cannot currently recall, is unclear. While some researchers do find differences in FOK accuracy between young and older adults (Perrotin et al., 2006; Perrotin et al., 2008; Souchay & Isingrini, 2004a; Souchay et al., 2007), others do not (Eakin & Hertzog, 2006, 2012a; Hertzog, Fulton, Sinclair, & Dunlosky, 2014; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002).

One possibility for such differences in empirical results is the task that is used to elicit metamemory judgments. Here, I outline a study that examines memory and metamemory performance in older adults using a specific memory phenomenon called *output interference* (OI), where memory performance gradually declines as a function of an item's position in a testing sequence, specifically for items that share semantic relationships (M. C. Anderson & Neely, 1996; Tulving & Arbuckle, 1963). This effect has its own precedent in the extant literature, including aging and memory (Smith, 1971, 1975), but it has yet to be connected to the topics of FOKs and metamemory in older adults. This study will not only directly examine whether there are age-related differences in FOK accuracy during output interference states, but it will also inform theories regarding the exact processes that underlie OI.

The remainder of this chapter will explore critical discrepancies in the extant metamemory research, where the aging literature has previously addressed interference effects with respect to memory judgments, and why OI might represent a new, rich environment for understanding memory awareness across the lifespan.

1.1 Aging & Metamemory

Metamemory refers to the ability to accurately review (*monitor*) and change the state (*control*) of one's own memory functioning (T. O. Nelson & Narens, 1990). Assessing the metacognitive abilities of older adults is extremely important, who are likely to experience heavy penalties for misjudging their own memory abilities, such as forgetting to take daily medication (Zogg, Woods, Sauceda, Wiebe, & Simoni, 2012). Thus, two goals of memory and aging research are to a) determine when and where memory prediction errors might occur in older adults, and b) explore the factors that contribute to age-related differences in metamemory performance where they are observed.

1.1.1 Initial Investigations

Early research in this field operationalized this ability by comparing the number or percentage of items that participants predicted they would remember (a "global" judgment of learning, or JOL) to the amount that they actually remembered. The general finding in this line of research is that older adults significantly over-estimate their memory abilities (i.e. predict remembering more items than they actually d0) compared to young adults, whose estimations are close to true performance (e.g. Perlmutter, 1978). In a study performed by Bruce, Coyne, and Botwinick (1982), for example, older adults (ages 60 - 79) over-estimated the number of items they would freely recall at test (out of 20) by an average of 2. This pattern was replicated in several additional studies (e.g. Coyne, 1985; Devolder et al., 1990), leading to the conclusion that memory prediction accuracy declines over the lifespan.

Despite the multiple replications, a number of inconsistencies in these results challenged this general conclusion of age-related metamemory deficits. First, several studies found no significant differences between young and older adults' global predictions, despite differences in true memory performance (Bruce et al., 1982). Second, several studies found the reverse pattern, i.e. better global JOL accuracy in older adults (Hertzog, Dixon, & Hultsch, 1990). This led to a critical review by Connor, Dunlosky, and Hertzog (1997) admonishing the use of global (or *absolute*) accuracy measures in favor of item-level (or *relative*) accuracy measures, such as the Goodman-Kruskal gamma correlation (L. A. Goodman & Kruskal, 1963; T. O. Nelson, 1984). The authors found that JOL accuracy, as measured by item-level correlations, is invariant between young and older adults and argued that the age-related differences found in global judgments can be attributed to the reliance of a mid-point anchor. The results of Connor et al. were instrumental in demonstrating that examining sensitivity to item-level variations in a learning sequence is critical to broad interpretations of metamemory accuracy across the lifespan.

This work on aging and JOLs sets an important precedent for the current proposal for a few different reasons. First, the work by Connor et al. (1997) and subsequent studies (Hertzog, Dunlosky, & Sinclair, 2010; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Hines, Hertzog, & Touron, 2015; Robinson, Hertzog, & Dunlosky, 2006) established that older and young adults generally do not differ with respect to item-level JOL sensitivity, with the exception of certain situations, such as when older adults overly-rely on familiarity (Daniels, Toth, & Hertzog, 2009; Toth, Daniels, & Solinger, 2011). More importantly, these studies demonstrated a shift from examining metamemory predictions as reflective of access to a quantity of information ("pure-accessibility") to examining them as the integration of multiple diagnostic cues that are present during the encoding and retrieval processes themselves (“cue-utilization“; Hertzog & Curley, 2018; Koriat, 1997; Robinson et al., 2006).

1.1.2 Origins of Feelings-of-knowing

The *feeling-of-knowing* (FOK)—or a prediction about future memory accuracy for a currently unretrievable item—has also been an important diagnostic tool for metamemory in adult learners across the lifespan. Unlike the JOL, which provides a subjective appraisal of confidence during memory encoding (c.f. Rhodes, 2016), FOKs indirectly examine the extent to which an individual has access to a particular piece of information in LTM. FOKs are maximally informative when examining judgments for items that an individual cannot recall, but “feels“ that the item can be accessed via recognition (T. O. Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982). The canonical procedure for eliciting FOKs is the recall-judgment-recognition (RJR) paradigm (Hart, 1965), where participants first learn a set of items with the instruction to remember them for a test later on and then engage in a cued-recall task after a brief retention period. During recall, participants are asked to rate their confidence that they would be able to correctly recognize the target item if it were shown to them (FOK). After providing a judgment after each recall trial, participants engage in a recognition memory task in which they must correctly recognize the items they studied at the beginning of the experiment. The degree to which item-level judgments for *unrecalled* items correspond to future recognition accuracy, typically measured using the Goodman-Kruskal gamma correlation (L. A. Goodman & Kruskal, 1963; T. O. Nelson, 1984), is broadly interpreted as the level of awareness an individual has of her access to information in LTM (T. O. Nelson & Narens, 1990). Thus, FOKs can be used to draw inferences about stability/changes in this awareness across the lifespan and whether older adults are less able to accurately monitor partial access to memory (Hertzog & Curley, 2018).

While the FOK phenomenon provides unique insight into specific metamemorial abilities, several factors make these judgments more difficult to interpret than JOLs. For one, the standard FOK does not have a direct analog by which to compare changes in access to criterial informaton. JOLs, by contrast, can be given either immediately after studying an item ("immediate JOL", or iJOL), which yields moderate metacognitive accuracy, or after studying all items in an intermediary testing session ("delayed JOL", or dJOL), which yields high metacognitive accuracy (T. O. Nelson & Dunlosky, 1991). The similarities in judgment conditions between the two types of JOLs allow researchers to directly contrast the conditions and heuristics that give rise to each. For example, the differences in judgment accuracy between immediate and delayed JOLs can signal a change in reliance on heuristics based on information in a short-term memory store to a reliance on information in a long-term memory store that closely matches the retrieval conditions in the final recall test (i.e. the *monitoring-dual-memories* account, T. O. Nelson & Dunlosky, 1991; although see Narens, Nelson, & Scheck, 2008, for a review of alternative accounts of the dJOL effect). FOKs do not have a close analog; instead, all empirical evidence regarding the nature of these judgments must be based on systematic manipulations of the stimuli and experimental conditions in the FOK task.

A second, but related, issue is the wide range of heuristics that potentially underlie the FOK construction process. Early researchers focused their conclusions on the correspondence between the *amount* of indirect information that participants are able to access at the time of judgment (e.g. a "trace-strength" account; Schacter, 1983) and the reported judgments; however, this hypothesis evolved to suggest that FOKs are constructed on the basis of conscious and nonconscious influences arising from the target search itself (Metcalfe, 2000). (Succinctly stated as "FOKs can access only the products of cognition"; Hertzog, Dunlosky, & Sinclair, 2010, p. 772.) One popular extension of this account is the *cue-familiarity hypothesis*, which posits that FOKs are constructed on the basis of the familiarity of the cue at the time of retrieval/judgment (Metcalfe et al., 1993). Reder and Ritter (1992), for example, demonstrated that FOKs can be almost entirely based on the familiarity of a cue by increasing the number of times that an arithmetic problem, but not its answer, is presented during study. A less stringent view of the importance of cue familarity is Koriat's *cue-accessibility* account (also referred to as the *partial-retrieval hypothesis*), where a cued item initiates the search process for the target and information spawning from that process is integrated into a judgment, regardless of whether that information is valid or not (Koriat, 1993; Koriat & Levy-Sadot, 2001). In recent studies, researchers have adopted a more holistic interpretation of FOK construction, one that hypothesizes that learners weigh multiple cues simultaneously (Hertzog, Dunlosky, & Sinclair, 2010; Hertzog et al., 2014). Along with the familiarity and accessibility of cue words, learners weigh the availability of information regarding the original study context, such as the emotional valence of cue-target pairs (A. K. Thomas, Bulevich, & Dubois, 2011) or mediators between the cue and target (Hertzog et al., 2014), often with significantly improved accuracy compared to controls. Information about the encoding context provides valid, albeit indirect, information by which to draw inferences during FOK construction, a phenomenon that is referred to as *non-criteral recollection* (Brewer et al., 2010; Yonelinas, 2002). Non-criterial recollection has been shown to improve FOK accuracy in both young and older adults (Hertzog, Curley, Castro, et al., 2020; Hertzog, Fulton, & Dunlosky, 2020; Hertzog et al., 2014).

The Role of Implicit and Explicit Information

There are several hypotheses related to the specific memorial cues are used to produce FOKs; for example, one popular account by Koriat (1997) segregates metacognitive cues into three categories: *Intrinsic*, *extrinsic*, and *mnemonic* factors. Intrinsic factors are those that reflect stimulus materials, such as the relatedness of a pair, while extrinsic factors are those that reflect the conditions of learning, such as study duration (Kelley & Jacoby, 1996). In contrast, mnemonic factors reflect information about items that are based on *subjective* inferences about memory, such as encoding fluency (Hertzog et al., 2003; Koriat & Ma'ayan, 2005).

The line of research that is most relevant to this study examines the role of *implicit* and *explicit* information, where implicit information refers to unintentional influences of previous experience on memory retrieval and explicit information refers to the influence of ideas that are intentionally encoded (D. L. Nelson et al., 2013; D. L. Nelson et al., 1992; D. L. Nelson & Zhang, 2000). During memory retrieval, both implicit and explicit information can be used to construct metamemory judgments.

Schreiber and Nelson examined the influence of information available during retrieval on episodic FOKs and predictions-of-knowing (POKs) in two experiments that manipulated the number of semantic associates (i.e. set size) for cue words (Schreiber & Nelson, 1998) and for target words (Schreiber, 1998). Here, the authors directly manipulated the role of implicit information using variable cue set-sizes: For cues with larger set sizes, for instance, implicit information will be influenced by the large number of associates that results from the retrieval process. In formulating the hypothesis, the authors distinguish between two potential retrieval accounts that might affect FOKs. The first, called the *partial-retrieval hypothesis* (Koriat, 1993), posits that FOK judgments are constructed on the basis of related information that comes to mind during retrieval and not on explicit access to criterial information. In the case of set-size manipulations, this account would hold in situations where participants give higher FOKs for items with a larger number of semantic neighbors than for items with smaller numbers of semantic neighbors. In this case, learners' judgments would reflect the feeling of greater access to information due to competition, regardless of whether the information is truly indicative of future memory performance or not. The alternative account, called the *competition hypothesis*, states the opposite pattern: Learners should report higher FOKs for items with fewer semantic neighbors compared to items with a larger number of semantic associates. Here, learners' judgments would reflect sensitivity to competition between items. In these two studies, Schreiber and Nelson found that FOKs were negatively correlated with the amount of competition during retrieval, such that higher FOKs were given to words with smaller set sizes (i.e. lower semantic competition) and lower FOKs were given to words with larger set sizes (i.e. higher semantic competition), supporting a competition hypothesis of FOK construction (Schreiber, 1998; Schreiber & Nelson, 1998). Schreiber and Nelson's work on the competition hypothesis was instrumental in later examinations of FOKs and implicit interference (Eakin & Hertzog, 2006, 2012a, 2012b).

Taken together, the results of these studies indicate that FOK accuracy in young adults is highest when learners have access to information that is diagnostic of (though does not necessarily directly point to) to-be-remembered items, such as memories based on recollection rather than familiarity (Hicks & Marsh, 2002), and lowest when direct information is unavailable or when implicit memory influences provide misleading information (Schreiber, 1998; Schreiber & Nelson, 1998). This is also the case for FOK judgment accuracy in older adults (e.g. Hertzog, Dunlosky, & Sinclair, 2010; recollection vs. familiarity, MacLaverty & Hertzog, 2009; Souchay et al., 2007); however, as I will discuss in the next section, delineations between different cue weightings become paramount in understanding decline and stability in FOK accuracy in later adulthood.

1.1.3 Sources of Age Differences in Episodic FOKs

An important consideration when examining feelings-of-knowing is the *type of memory* learners are asked to make judgments about. Here, the distinction between *semantic* and *episodic* FOKs not only delineates functional differences in the task, but also differences in common findings regarding aging and metamemory. The accuracy of semantic FOKs, which are FOKs about facts or knowledge, remains stable over the lifespan, with very few exceptions (e.g., Allen-Burge & Storandt, 2000; Eakin, Hertzog, & Harris, 2014; Souchay, Isingrini, & Espagnet, 2000; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007; for a review of earlier studies, see Hertzog & Hultsch, 2000). The common explanation for this is that semantic stimuli involve topics that are familiar to adults and information that has been previously mastered (Hertzog & Curley, 2018). For example, semantic FOK tasks may ask participants to recall the name of the U.S. Vice President during Jimmy Carter's administration.[1] While an older adult may not be able to directly recall the name, she might be able to remember certain details about the former VP, such as being from the Midwest and involved in investigating national intelligence organization abuses,[2] and use that information to form an accurate FOK. While spreading activation from some items may prevent older adults from accessing relevant information, enough items are sufficiently recollected in order to attain above-chance FOK accuracy at the level of young adults.

In contrast, the stability of episodic FOK accuracy over the lifespan is still under debate. Early studies indicated no significant differences in FOK accuracy between young and older adults either for scale (Lachman, Lachman, & Thronesbery, 1979) or binary/relative (Butterfield, Nelson, & Peck, 1988) FOK judgments. However, a paper by Perfect and Stollery (1993) challenged these findings by demonstrating that significant age-related differences in memory appraisals are evident when age changes in episodic memory prevent older adults from accessing diagnostic cues. In more recent years, a number of conflicting studies have examined FOK accuracy in further detail with differences in their findings. These studies are outlined below and classified by two broad, but aptly-named schools of thought: Experiments that support an *inferential deficit hypothesis*, i.e. that age-related differences in FOK accuracy are due to diminished decision support systems in older adults, and those that support a *memory-constraint hypothesis*, i.e. that age-related differences in FOK accuracy are simply attributable to lower episodic memory strength in older adults (Hertzog, Dunlosky, & Sinclair, 2010).

Inferential Deficits via Neuropsychological Factors

The most prominent evidence of age-related differences in episodic FOK accuracy is from research conducted by a group from the Université de Tours in France. In their

[1] Answer: Walter Mondale.

[2] In 1975, Mondale was part of a Senate committee that investigated abuses in the CIA, NSA, and FBI. It was informally known as the "Church Committee" after its chair, Senator Frank Church.

first study on the subject, Souchay, Isingrini, and Espagnet (2000) asked young and older adults of good physical and mental health to learn and give binary (i.e. “yes”/“no”) FOKs for 36 moderately-associated French noun pairs using the common recall-judge-recognize paradigm (Hart, 1965; Schacter, 1983). While young adults showed moderate item-level relationships between FOKs and later recognition (i.e. γ = 0.40), older adults’ judgments were no more accurate than chance (i.e. γ = -0.06). Several other studies from this group have replicated this general pattern of age-related differences in episodic, but not semantic, FOK accuracy (Morson, Moulin, & Souchay, 2015; Perrotin, Isingrini, Souchay, Clarys, & Taconnat, 2006; Perrotin, Tournelle, & Isingrini, 2008; Sacher, Isingrini, & Taconnat, 2013; Sacher, Landré, & Taconnat, 2015; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007), with some connections to other metacognitive influences, such as control of study (Souchay & Isingrini, 2004b), decisions made during study (Souchay & Isingrini, 2004a), and the role of recollection (Souchay et al., 2007).

The primary conclusion from these studies is that age-related deficits in FOK accuracy arise from deficiencies in executive control processes, which themselves arise from declines in frontal and medial temporal cortical areas. The authors connected these studies to earlier research by Shimamura and Squire (1986) and Janowsky, Shimamura, and Squire (1989), who demonstrated that patients with damage to the frontal areas of the brain, such as those with Korsakoff syndrome or with frontal lobe lesions, show diminished FOK accuracy. When also considering accounts of diminished frontal lobe volume and functioning in older adults (Dempster, 1992; Moscovitch & Winocur, 1992), one can make the reasonable conclusion that metamemory accuracy will decline in older age, concurrent with decreases in frontal lobe functioning. Indeed, these studies relate age differences in episodic FOK accuracy to declines in executive functioning, such as through significant partial correlations between judgment accuracy and neuropsychological measures (Perrotin et al., 2006; Perrotin et al., 2008; Souchay et al., 2000); however, a direct link has yet to be made to physical indicators of frontal lobe deficiencies (e.g., via neuroimaging) in older adults.

Memory Constraints via Diminished Access

The studies carried out by the Université de Tours research group, which largely support a inferential-deficit hypothesis, have yielded some discrepant conclusions. The role of frontal lobe functioning in general episodic memory performance is the most prominent example, where the authors argue that age-related declines in executive functioning are more closely related to inaccuracies in metamemory judgments than to general declines in episodic memory (i.e. Craik et al., 1990; Rosen et al., 2002). Sacher et al. (2015) specifically addressed this issue by using signal-detection techniques to ascertain the separate influences of overall memory performance and metamemory ability in age-related FOK accuracy differences. The authors used a Brier score analysis in order to examine these two influences and concluded that memory-independent processes have a significant effect on older adults' FOK judgments, even after accounting for overall memory performance. Two issues call this argument into question, however: First, Sacher et al. do not directly compare age groups using equated memory performance in order to validate the Brier score analysis. Second, the authors argue that SDT measures, such as the meta-d' statistic proposed by Maniscalco and Lau (2012), can estimate the separate effects of metamemory ability and underlying episodic memory content, regardless of differences in underlying memory performance. However, such inferences are predicated on the assumption that the quality of metacognitive judgments can be normalized by stimulus sensitivity (Fleming & Lau, 2014). This assumption is well-suited for tasks with judgments that follow from first-order decisions, such as sensory and post-recognition confidence judgments, but may not be appropriate for higher-order metacognitive judgments that are constructed on the basis of subjective heuristics evolving from the retrieval attempt. Indeed, Mazancieux, Dinze, Souchay, and Moulin (2020) hint at this possibility in a secondary analysis in which metacognitive sensitivity in FOKs is shown to be significantly related to recall performance, unlike RCJs, where metacognitive sensitivity was not related to recognition memory performance.

An alternative explanation is that age-related differences in FOK accuracy stem from differences in episodic memory, where young adults, who recall more items than older adults in equal retention periods, are basing their judgments off of stronger memory traces. Put another way, a memory-constraint hypothesis suggests that a lack of availability to diagnostic cues is a function of age-related declines in episodic memory (Perfect & Stollery, 1993). Hertzog, Dunlosky, and Sinclair (2010) directly test this account in a number of different ways. They manipulated the quality of memory representations by varying the number of item presentations (i.e. 1, 2, or 4 times) for both young and older adults and demonstrate that FOK accuracy increases as a function of repetitions (i.e. memory strength). Importantly, Hertzog, Dunlosky, and Sinclair introduced differential delays between study and test in order to equate memory performance between young and older adults. Recall memory performance was equated for young adults with a one-week delay between study and test and older adults with a 2-day delay between sessions for all repetition conditions (2010). Overall, the results indicate that there are no significant effects of age in FOK accuracy when equating underlying memory strength; thus, the lack of access to information is a strong influence on episodic FOKs. The Hertzog group (which I am a part of) confirmed this pattern in a recent replication (Hertzog, Curley, Castro, & Dunlosky, 2020).

Repairing Accuracy Deficits

While the memory-constraint hypothesis provides a parsimonious explanation for extant experimental findings, the account is complicated by the fact that FOK accuracy equivalence between age groups is also found in studies that do not equate memory performance between young and older adults. For example, MacLaverty and Hertzog (2009) did not detect any significant differences in episodic FOK accuracy between age groups, despite the fact that young adults correctly recalled 30% more target words on average than older adults. Similarly, a study by Eakin, Hertzog, and Harris (2014) found no age-related differences in episodic FOK accuracy for name-face pairs, despite a significant difference in recall performance for episodic stimuli ($M_{YA} = 0.10$, $M_{OA} = 0.03$).

Given this, a related and more important issue—both for metamemory and aging as well as the current study—is the availability of *diagnostic* information during FOK construction. This theoretical stance integrates the general memory-constraint (Hertzog, Dunlosky, & Sinclair, 2010) and multiple cue-integration (Hertzog et al., 2014) accounts by postulating that all learners, both young and old, can provide equally-accurate FOKs if they have access to cues that are diagnostic of future memory performance. A difference in memory trace strength due to age-related declines in episodic memory functioning (e.g., Sacher et al., 2015) is just one common path to diminished access to such cues; thus, age invariance in FOK accuracy is not strictly reliant on equating recall performance across age groups.

To date, the most powerful evidence for this view comes from studies examining FOK accuracy and *non-criterial recollection*, or the retrieval of information related to a target item that is diagnostic of future memory performance, but does not directly evoke retrieval of the target itself (Yonelinas & Jacoby, 1996). Non-criterial recollection has been specifically linked to familiarity deficits in older adults (Toth & Parks, 2006) as well as to cortical regions that are likely to degrade with age (Diana, Yonelinas, & Ranganath, 2007), making it a likely source of judgment accuracy in metamemory experiments. Accordingly, Hertzog et al. (2014) showed that retrieval of information related to the study environment (in this case, a sentence or image mediator connecting a cue word and target word pair) significantly improved gamma correlations between scale FOK judgments and recognition memory accuracy for unrecalled items in young adults. While the specific impact of non-criterial recollection in repairing age-related differences in FOK accuracy is under debate (e.g. A. K. Thomas et al., 2011), recent research from the Hertzog group has provided evidence in favor of the hypothesis that access to non-criterial information during recall increases FOK accuracy in both young and older adults (Hertzog, Curley, Castro, et al., 2020; Hertzog, Fulton, & Dunlosky, 2020).

Lack of Constraints/Deficits in Implicit Interference

While research on non-criterial recollection focuses on how to *increase* access to information that will improve metacognitive judgment accuracy, an equally compelling line of research (and one that is paramount to the proposed study) is understanding the factors that *decrease* access to informative cues in young and older adults. While this issue has had comparatively little coverage in the extant literature, the role of *interference*—a specific inhibitory mechanism that impairs one's ability to remember an item that was studied with similar items (M. C. Anderson & Neely, 1996)—has had a demonstrable role in FOK judgment construction in young adults. Metcalfe, Schwartz, and Joaquim (1993), for example, provide early experimental evidence for diminished metacognitive accuracy under interference conditions. The authors examined FOK and TOT judgments under several proactive interference (PI) experimental manipulations using cue and target repetition at study (i.e., A-B, A-B; A-B, A-B'; A-B, A-D designs). The authors found that these judgments were related to the number of presentations of the cues, but not the targets, such that FOK judgment magnitudes were significantly higher for items with repeated cues. These findings were interpreted with respect to metamemory judgment construction rather than awareness of PI; specifically, the authors argue that the results support FOK construction based on *cue-familiarity* rather than information gleaned from a target or even access to information overall (i.e. an *accessibility account*; Koriat, 1993). Maki (1999) developed this research further by studying the effects of retroactive interference (RI) on JOLs and FOKs, where learners reported significantly higher estimates for both types of judgments when stimuli were repeated and their responses were semantic associates. Maki interpreted these results as favoring the *competition account* of metamemory construction (Schreiber, 1998).

Importantly, Eakin and Hertzog published a set of studies examining metamemory under manipulations of retrieval interference (Eakin & Hertzog, 2006, 2012a, 2012b). In their first study, Eakin and Hertzog (2006) examined cued recall, FOK, and POK performance between young and older adults for items that have small or large set sizes and under intraor extra-list cueing. Similar to previous research, set-size effects were only eliminated in younger adults under intralist cueing. Importantly, metamemory judgment accuracy was equivalent across age groups, indicating that all participants demonstrated sensitivity to implicit interference effects. Differences in the relationship between POKs and recognition were significant, however; when considering all items, mean gamma correlations for both age groups were moderate ($\gamma \approx 0.4$), while mean gamma correlations for unrecalled items were practically zero. This indicates that interference effects largely influence recall, but not recognition memory, and that inaccuracies in memory judgments were related these interference effects during recall.

The authors also published two additional studies examining the influence of implicit interference on FOKs in young and older adults (Eakin & Hertzog, 2012a) and immediate JOLs in young adults only (Eakin & Hertzog, 2012b) using similar methods. Along with replicating the set-size effects in cued recall, the researchers found that both young and older adults' FOKs reliably tracked recall, but not recognition (2012a), and that iJOLs were not diagnostic of future recall performance (2012b). These studies provide further evidence for the hypothesis that interference states are localized to retrieval during cued recall and that only judgments closely related to the retrieval interference experience itself are reliably diagnostic of recall performance. These studies by Eakin and Hertzog (2006, 2012a, 2012b) are some of the primary influences on the proposed project.

1.2 Memory Interference

The most surprising outcome of Eakin and Hertzog's studies on aging and FOKs using implicit interference (2006, 2012a) is a lack of age differences in overall judgment accuracy, despite the clear negative effects of extralist cueing and large set sizes. Further, the results of the cued recall task indicate that older adults are particularly sensitive to interference effects during retrieval. In Eakin and Hertzog (2006), for example, cued recall exhibited a three-way interaction between cue set size, cueing procedure, and age group, such that set size effects (i.e. lower recall for items with larger set sizes) were eliminated under intralist cueing for young adults, but not for older adults. These results are inconsistent with Schreiber and Nelson's (Schreiber, 1998; Schreiber & Nelson, 1998) competition hypothesis, which predicts that FOKs should be negatively correlated with competition. In the case of Eakin and Hertzog (2006), FOK accuracy for older adults, who show greater effects of interference during recall, should be significantly lower in the conditions that encourage implicit interference. Given these results, one could conclude that memory interference cannot account for differences in FOK accuracy between young and older adults, despite significant differences in underlying memory performance.

An alternative hypothesis is that different interference paradigms yield different metamemory outcomes. While implicit (Eakin & Hertzog, 2006, 2012a) and proactive (Diaz & Benjamin, 2011; Maki, 1999) interference paradigms exhibit similar patterns of metamemory performance (which I will explore in more detail later in this section), they also share many fundamental properties that make similarities in these results probable. At the end of this section, I will introduce an older, but potent interference paradigm, *output interference*, as a potential source of FOK inaccuracy in older adults that differs significantly from the interference paradigms previously used.

1.2.1 Proactive Interference

Proactive interference (PI) refers to the deleterious impact of irrelevant information learned prior to the main encoding trials on memory for relevant targets (M. C. Anderson & Neely, 1996). In standard PI paradigms, participants initially study a set of irrelevant items ("List 1"), followed by the true set of cue-target pairs ("List 2"), and ending with either a cued-recall or recognition memory test on the relevant items ("List 2"). Studying List 1 items interferes with the recollection of List 2 items, resulting in decreased memory for relevant items, particularly for cued recall and longer retention periods (Postman, Stark, & Fraser, 1968).

PI is a particularly important topic in aging research for many different reasons. For one, PI has been argued to be the major source of forgetting in everyday life (Underwood, 1957), where years of prior learning can interfere with memory for new information. Additionally, PI has been shown to have greater effects on older adults, potentially reflecting decreases in the ability to inhibit irrelevant information in both long term and working memory tasks (Lustig, May, & Hasher, 2001). Indeed, leading theoretical accounts hypothesize that increases in PI are directly related to increases in search set evoked by the cue word (M. C. Anderson et al., 1994; M. C. Anderson & Neely, 1996; Watkins & Watkins, 1975; Wixted & Rohrer, 1993).

Effects of PI on Metamemory

This interference paradigm has also been used in metacognitive research to help differentiate between accessibility (Koriat, 1993) and competition (Schreiber, 1998; Schreiber & Nelson, 1998) accounts of metamemory construction. A study by Maki (1999) was one of the first to directly compare these two accounts of metamemory construction for both FOKs and JOLs. Learners reported significantly higher estimates for both types of judgments when stimuli were repeated and their responses were semantic associates. These results were not limited to manipulations that only involve cue words, leading to the conclusion that a more general metamemory judgment construction mechanism than cue-familiarity or accessibility to the target is plausible, given interference at test. Thus, the results of Maki's study are consistent with Schreiber's (1998) competition account of metamemory judgment construction (Maki, 1999).

Research by Wahlheim (2011) provides further support for a competition account of metamemory using a PI framework. Here, the author replicated the typical delayed JOL effect, i.e. higher accuracy for dJOLs than iJOLs, but still found that reported dJOLs were higher in magnitude under interference conditions. This inflation in dJOLs was attributed to high-confidence judgments given to intrusion errors during recall. The implications of these data are two-fold: First, dJOLs are susceptible to interference during retrieval, even though these judgments rely on more "valid" information (i.e. traces from LTM) than do iJOLs (i.e. noise from STM). Secondly, the effects of interference during the judgment process leads learners to become over-confident. This finding challenges the conventional monitoring-dual-memories (MDM; T. O. Nelson & Dunlosky, 1991) account, which postulates that having access to information in LTM should increase judgment accuracy, regardless of the retrieval context. The conclusions support a theory of judgment construction that emphasizes the role of the retrieval process in dictating the quality of information used for a given judgment.

More recent research on metamemory and PI challenges the notion that attenuation in accuracy estimates due to interference result from how items are themselves processed. Diaz and Benjamin (2011) studied JOLs in conditions of PI and "release" from PI (Wickens, 1973), where some cues were repeated in a block followed by novel ones. Overall JOLs did decrease across trials in which PI was built up, but they continued to decrease over trials, even after new cues were used. The judgments in the latter case did not follow the increase in memory performance after new cues were presented. JOLs also decreased equally for pairs with novel cues as well as those with repeated cues, even though recall for pairs with novel cues did not decrease over trials. The authors argue for an account of metacognition in which learners have a global, but not item-specific, awareness of interference.

Phenomenology of PI

The most curious aspect of interference research is that there are few studies that attempt to confirm that what memory theorists and learners call "interference" is congruent with mechanistic accounts of memory interference. (Tulving, 1989, expertly coined this as the "doctrine of concordance", or the oft-untested assumption in cognitive psychology that memory functioning and experience are the same.) Metamemory research on PI, such as that from Maki (1999), is clear that learners do not use mnemonic cues to gauge the effects of interference and, instead, rely on what Diaz and Benjamin (2011, p. 202) refer to as a "naive theory of memory"; however, these interpretations are only sufficient in partially describing how learners integrate information to make memory judgments during PI and not the experience of PI itself.

1.2.2 Implicit Interference

Interference research distinguishes between two broad categories based on the amount of awareness learners have during the task: In *explicit* memory interference tasks, such as PI, learners are generally aware of sources of disruption, e.g. memory for List A when attempting to recall List B. In *implicit* interference tasks, however, learners are not consciously aware of such disruptive sources. An example relevant to this research is implicitly activating activated associates of the cue during recall (Eakin & Hertzog, 2006).

To be clear, I am careful to distinguish between interference paradigms that rely on *implicitly-activated* information and interference paradigms that examine *implicit memory* itself. The former describes any memory task that involves activation of competing information that a participant is not explicitly aware of, while the latter describes interference during implicit memory tasks, such as repetition priming and stem completion tasks (Roediger, 1990; Schacter, 1987). This paper will not go into detail about interference during implicit memory tasks, although it is worth mentioning that older adults are more susceptible to interference effects in these tasks than young adults (Ikier & Hasher, 2006; Ikier, Yang, & Hasher, 2008).

Many important demonstrations of interference from implicit sources come from Douglas Nelson and colleagues (D. L. Nelson & McEvoy, 1979; D. L. Nelson, McEvoy, & Schreiber, 1990; D. L. Nelson, McKinney, Gee, & Janczura, 1998). These studies demonstrate that *cue-set-size*, or the number of associates that are implicitly co-activated when a cue is shown at test, is an important determinant of memory performance. Specifically, the probability of recalling an item is negatively correlated with the cue set size, i.e. targets that are associated with cues that have a large number of semantic associates are less likely to be recalled. One plausible explanation for Nelson et al.'s the cue-set-size effect is that interference occurs when semantic associates to a cue word are co-activated during retrieval, creating a negative implicit effect on memory performance. This is best explained using the Processing Implicit and Explicit Representations model (PIER2; D. L. Nelson et al., 1998), which holds that processing a cue activates semantic associates implicitly. The level of co-activation is thought to be positively correlated with the amount of interference that learners experience during retrieval, although cue-set-size effects can be eliminated when a cue and target with shared associates are studied together (i.e. intralist cueing), which reduces the number of potential associates to those that are shared by the word pair (D. L. Nelson & McEvoy, 1979; D. L. Nelson et al., 1990).

Older adults are thought to be particularly prone to cue-set-size effects. McEvoy, Holley, and Nelson (1995), for example, found significant age-related differences in recall performance in an extralist cueing paradigm, or a task in which the target is cued by a previously-unseen word. The authors interpret this with respect to an inhibitory-deficit account (Hasher & Zacks, 1988), where older adults are less able to inhibit the influence of co-activated associates during retrieval. This interpretation is consistent with later research suggesting that older adults have greater difficulty discriminating between targets and competitors without memory training (e.g. Badham et al., 2016).

1.2.3 Output Interference

A central question to the current project is whether all types of interference give rise to the same memorial experiences. If they do, then we would expect any interference paradigm would give rise to competition due to co-activated associates, regardless of the manner in which the associates are activated (i.e. implicitly or explicitly). This general competition hypothesis (D. L. Nelson et al., 1998; Schreiber, 1998; Schreiber & Nelson,

1998) would also be expected to yield metamemory judgments that decrease in magnitude with larger set sizes (Eakin & Hertzog, 2006, 2012a, 2012b). However, several studies have indicated that some interference states are qualitatively different from ones such as proactive and implicit interference. *Output interference* (OI), or the gradual decrease in retrieval as a function of an item's position in a testing sequence, is one such example, which "violate[s] the widely held idea... that interference is initiated by competition for a shared retrieval cue" (M. C. Anderson & Neely, 1996, p. 270).

What follows is a brief overview of important research and theories regarding OI effects, as well as computational accounts of this interference effect. Importantly, this section will also examine how OI differs from other interference paradigms and why the phenomenological experience of OI might give rise to metacognitive judgments patterns that differ from general competitions accounts.

Early Research

Original theories of output interference suggested that the effect is agnostic of item-type, and that any memory task with sequential retrieval trials will exhibit decreases in accuracy with serial position. Tulving and colleagues (Tulving & Arbuckle, 1963, 1966) demonstrated this effect when using a cued recall task with simple noun-number paired associates. The authors concluded that decreases in response accuracy were indicative of a loss of information in short-term memory between study and test, particularly for items that were tested at the end of the recall sequence. However, these interference effects have been shown to occur in tasks that control for degradations in a short-term memory store. Smith (1971), for example, demonstrated that giving learners a task in between study and test to occupy their short-term memory does not eliminate output interference effects (although the criterion task differed from those from Tulving and colleagues' studies, i.e. recognition). Participants showed decreased free-recall accuracy for categorized words across sequential test trials, despite an interpolation task between learning and retrieval

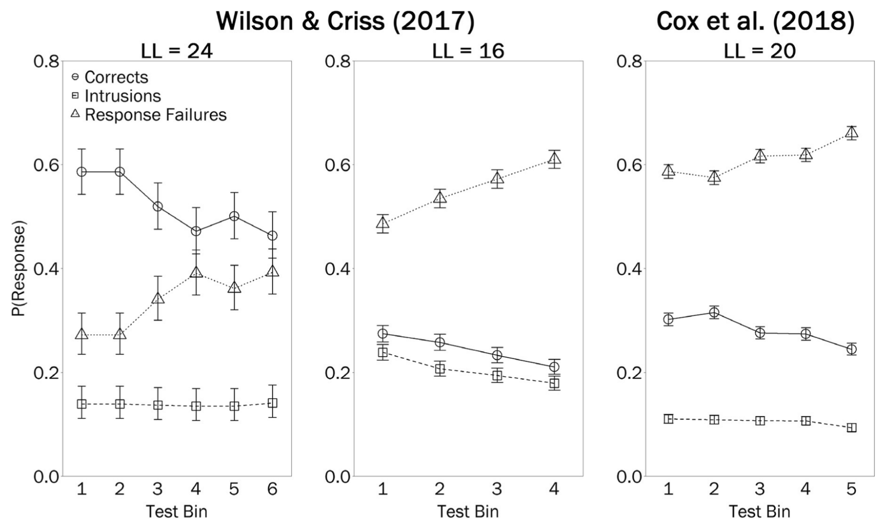

Figure 1.1: An overview of typical output interference effects in cued recall. Typical OI effects include decreased accuracy rates, increased omission rates, and relatively stable commission error rates. From Wilson et al. (2020).

procedures. Output interference must therefore be a function of trace activation in LTM, resulting as a consequence of repeated retrieval attempts.

Another early hypothesis regarding output interference is that the effect only exists when targets have a shared cue. Generalizations of memory interference experimental paradigms challenge this notion. In the study by Smith (1971), the study items consisted of 49 items, or 7 exemplars from 7 taxonomic categories. In a similar study by Roediger and Schmidt (1980), learners studied lists of exemplars from taxonomic categories and were cued either with only the category label or with the category label and 4 exemplars and asked to recall the words that were presented during study. The results indicate that output interference effects persist across these different conditions, such that recall accuracy by serial position coefficients are similar between conditions in which learners are cued with just the category label or the category label plus 4 exemplars, conditions with differing numbers of to-be-remembered exemplars, and conditions in which the categories in the study list are related versus unrelated. The results of these studies suggest, then, that interference effects are not dependent upon either cue or category ("cue-independence").

Current theoretical views of output interference do not postulate a generalized competition account; rather, the extant literature suggests that, while output interference is not specifically dependent upon shared retrieval cues, it is dependent upon the type of information being accessed during retrieval. Neely et al. (1983) provided an early demonstration of this using a procedure modified from that of Roediger and Schmidt (1980). Participants were asked to study lists of 5 category exemplars and then to choose the previously-seen targets during a speeded yes/no recognition test. At test, the researchers manipulated implicit influences on memory by presenting categorically-related lures ("primes") prior to certain recognition trials. The results show two distinct patterns. First, showing a related prime prior to a yes/no recognition decreased response latencies compared to preceding the test trial with an unrelated word, indicating that increasing semantic activation for an item facilitates memory retrieval. Second, and most importantly, recognition response latencies for items that were preceded by 6 related primes were significantly higher than those for items preceded by only 2 primes. This second set of results has been key to the conception that retrieval is a source of interference, and that accessing previously-encoded information facilitates increased error rates across trials (M. C. Anderson & Neely, 1996; Criss et al., 2011; Wilson et al., 2020).

Mechanisms of Output Interference

Recent empirical investigations regarding the mechanistic properties behind OI have focused on modeling changes during memory retrieval. Simulations using the search of associative memory (SAM; Raaijmakers & Shiffrin, 1981) model, where memory traces are activated by co-occurring features during retrieval, have proven to be informative of this issue. Specifically, SAM postulates that the process of retrieving an item carries the additional benefit of encoding extra information about the item, known as *learning during retrieval* (e.g., Carrier & Pashler, 1992; Roediger & Karpicke, 2006). For items that are successfully recovered, SAM engages in a process called *incrementing* in which the strength of the associations between a cue, the learning context, and target are strengthened. Incrementing was instrumental in early demonstrations of interference effects in free recall (Raaijmakers & Shiffrin, 1981) and later investigations using the retrieving effectively from memory (REM; Shiffrin & Steyvers, 1997) model to investigate OI effects in recognition memory (Criss et al., 2011; Koop et al., 2015).

A key component to retrieval processes in these simulations is one that limits the number of retrieval attempts per memory test trial. The SAM model originally used this component, referred to as a *retrieval filter*, simply as an economical device to prevent the model from continuously engaging in retrieval search. This mechanism was adapted from a computational model of interference in free recall by Rundus (1973), where the number of unsuccessful retrieval attempts is bounded by the integer parameter m and re-sampling of previously-retrieved (and recently-activated) items causes an individual to terminate search earlier.

The concept of the retrieval filter continues to be an important tool for understanding retrieval mechanisms in free recall, cued recall, and recognition (Wilson et al., 2020), despite its innocent origins. For example, The Hypothesis Generation (HyGene) model (R. P. Thomas et al., 2008) implements this mechanism as T_{Max}, which corresponds to the maximum number of retrieval failures during the hypothesis comparison process, in order to explore the effects of WM constraints in subadditivity (Dougherty & Hunter, 2003a, 2003b).

The combination of the incrementing and retrieval failure mechanisms in formal models of episodic memory provides the basis for OI effects in cued recall: Previously-retrieved items have increased activations and are likely to be sampled in later trials with similar cue information (e.g., categorically-related), causing the retrieval failure count to reach its maximum earlier in the sampling process (Wilson et al., 2020). Thus, for consecutive recall trials in which the cues are related to each other, recall performance declines as a function of serial position due to interference from traces activated earlier in the testing sequence. An important point to note is that these computational accounts largely examine the degree to which patterns of memory performance that are simulated using these parameters fit true patterns of performance in OI tasks and do not specifically examine the role of competition during retrieval.

Aging and Output Interference

At the time of writing this proposal, there has been very little research conducted on OI effects in older adults, and certainly none from the past few decades (Kausler, 1994). However, indirect evidence from related tasks are concordant with the idea that older adults show greater interference effects at test. For example, Duchek (1984) demonstrated that elderly learners show retrieval deficits related to semantic context. Older and young adults were asked to engage in a semantic orienting task for paired associate learning in which they were asked about a target word's category membership. At test, individuals were prompted with either semantic or rhyming cues and asked to recall the target word. While both age groups demonstrated greater memory accuracy for items with semantic cues at test, young adults' memory for items with semantic cues was significantly greater than that for older adults. Ducheck ascribes this to a general deficit in older adults to reinstate specific semantic contexts at test, although the extent to which this finding is dependent upon competition during test that is influenced by semantic contexts is unclear (Kausler, 1994).

To date, only a few direct investigations regarding OI in older adults have been conducted. Taub and Walker (1970) initially provided evidence for age-related differences in interference effects. Learners were asked to recall a word from a previous list that overlaps with a current one before fully recalling the words that are unique to a current list—a procedure that is meant to induce interference for specific items in a list. Older adults demonstrated significantly lower memory accuracy overall as well as qualitatively different levels of recall over trials, which supports the hypothesis of an age difference in OI effects.

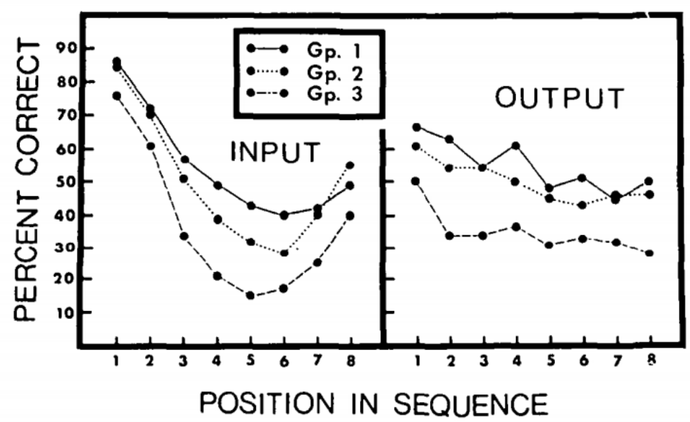

Figure 1.2: Results from Smith (1975) regarding recall performance for young ("Gp. 1"), middle ("Gp. 2"), and older ("Gp. 3") adults. Output interference effects for recall as a function of serial position at test ("OUTPUT") are invariant across age groups.

The most direct test of the effects of aging on OI by Smith (1975), however, does not support this general conclusion. In his study, Smith carefully examined the separate influences of input (STM) and output (LTM) interference by pairing study order with the subsequent test order in a factorial design. Participants from three age groups (20-39, 40-50, and 60-80 years of age) were asked to study 8 exemplars from 9 normative categories (Battig & Montague, 1969). While there were no significant trends related to input position, each of the 3 age groups demonstrated similar output interference effects during test, despite having significantly different levels of free recall performance (Figure 1.2). Additionally, rates of omission and commission errors were similar across the 3 age groups, suggesting a lack of qualitative differences in recall outcomes.

Phenomenology of Output Interference

A major discrepancy in recent OI literature is the lack of explanation for *how* OI occurs and its relation to actual subjective memory experiences. If OI is indeed a ubiquitous experience in everyday life (M. C. Anderson & Neely, 1996), then connecting the "feeling" of interference during OI tasks should provide converging evidence for the mechanisms that have been hypothesized to give rise to OI (e.g., Raaijmakers & Shiffrin, 1981; Rundus, 1973; Wilson et al., 2020). Unfortunately, there does not appear to be any such research in the extant literature.

The experimental paradigm that is most informative of how to connect the mechanisms and experiences of OI is *retrieval-induced forgetting* (RIF; M. C. Anderson et al., 1994). Here, participants study categorized lists of words that they will either see again ("Retrieval pracice", or *Rp*, items) or not ("No retrieval practice", or *Nrp*, items). Importantly, specific words within the Rp list are not shown again for restudy (*Rp-*), and memory for these items is examined against items that were designated to be shown again (*Rp+*). While the RIF paradigm is not the same as OI, the results are similar: Retrieving items that were restudied earlier (Rp+) decreases the probability of retrieving related, but once-seen items (Rp-). This general finding in RIF has been shown to be independent of the cue word (M. C. Anderson, Bjork, & Bjork, 2000) and number of exposures to items in the intial study phase (Hulbert, Shivde, & Anderson, 2012).

M. C. Anderson and Neely (1996) described two particular theoretical accounts of RIF that were influential in later research: One that implicates the roles of inhibition and competition (which I will refer to as the *competition* account), and a second that argues that recently-activated items "block" the retrieval of related items later in the retrieval sequence (which I will refer to as the *retrieval-suppression* account; c.f. Bäuml, 1998). Anderson and colleagues favor the former account and argue that RIF is a byproduct of spreading activation to competitors and the over-activation of related items that have been recently retrieved. These items compete for access and require a process to engage in a selective retrieval process that ignores highly-activated non-target competitors in favor of the more weakly-activated target. This is referred to as "response override" (M. C. Anderson & Levy, 2007, p. 81) and has been theorized to be directly related to executive functioning, i.e. lower inhibitory abilities result in failures to engage in a response override (M. C. Anderson, 2003; M. C. Anderson et al., 2000; M. C. Anderson & Neely, 1996; Hulbert et al., 2012).

The competition account cannot fully account for RIF and OI results, however. For one, interference effects are dependent upon the pre-experimental associative strengths of the cues and targets. In Anderson and colleagues' account, the source of interference arises soley from the build up of related items that increased in activation due to previous retrieval attempts. Here, the strengthening of the category cue to item associations during retrieval should produce increased interference from these items to other items with the same cue, regardless of the pre-experimental associative strength. This is not strictly the case, however; Bäuml (1998) demonstrated that retrieving target words with moderate normative associativeness to a cue word produced output interference only for items that were strongly associated with the cue word and not for those that were weakly associated with the cue. This was later replicated in a RIF study by Williams and Zacks (2001).

Importantly, the competition account has been argued to be insufficient because of a lack of specific evidence for the role of inhibition in RIF. Anderson and colleagues posit that retrieval interference occurs as a result of active inhibition of co-activated, but competing representations early in a retrieval sequence. Later in the sequence, when these items are no longer competitors to ignore, but are the correct targets, they continue to be downgraded as a result of this early inhibition. Experimental evidence does not support a role for active inhibition, though; in a study by Williams and Zacks (2001), for example, final memory accuracy for retrieval-practice items that were not studied (Rp-) was not statistically different from items that did not undergo retrieval practice at all (Nrp). If there was an active response override (M. C. Anderson & Levy, 2007), then the representations for Rp- items later in a retrieval sequence would be down-regulated as a result and would have a lower chance of being retrieved than Nrp items. Williams and Zacks (2001) concluded that RIF is a function of non-inhibitory processes.

Further evidence against a competition account comes from a study by Aslan et al. (2007), who examined RIF in older adults. The null hypothesis (i.e. a competition account) holds that older adults have difficulties preventing irrelevant information from entering into WM (Hasher et al., 1991; Hasher & Zacks, 1988; Hasher et al., 1999). If inhibitory processes have an active role in retrieval interference, then memory performance betwen older and young adults should be qualitatively different. However, similar to OI effects in the study by Smith (1975), there were no interaction effects to suggest that young and older adults had significantly different RIF results (although older adults recalled fewer items on average; Aslan et al., 2007). The authors argue that the results indicate that inhibitory deficits in older adults are task-dependent; however, given the reliability of the inhibitory-deficit account and the results of previous RIF studies, these results also argue against an inhibitory account of aging and RIF altogether.

In contrast to an inhibitory account of interference, the retrieval-suppression account of OI assumes that activation of previously-retrieved items *precludes* future target items from entering awareness rather than an intentional *exclusion* (i.e. inhibition) that occurred earlier in the retrieval sequence. The largest difference in this account is that learners are unable to recall the identity of a target item precluded from memory, whereas inhibited items can be recalled, but learners are unable to distinguish the context in which it was first studied. Anderson and Neely (1996) provide an example of metamemorial awareness during such an occlusion of a target item:

> Sometimes our ability to recall our current parking location seems blocked by the intrusion of similar episodes. When this occurs, we often feel confident that we know where we parked, but that recall of the location demands that we penetrate through memories that get in the way. (pp. 237-238)

This conception of interference is largely informed by early memory research on blocking (Wickelgren, 1976) and response suppression (Postman et al., 1968) in interference. Unfortunately, the type of evidence needed to empirically test a suppression/blocking account of interference is not available in a standard interference paradigm; therefore, studies that reference this hypothesis are limited to these early experiments.

1.3 Overview of the Current Experiment

Previous research has reliably demonstrated that buildup of competing semantic associates over successive cued recall attempts results in decreases in cued recall performance over successive trials (c.f. Tulving & Arbuckle, 1963, 1966). This has been demonstrated in recognition (e.g. M. C. Anderson & Neely, 1996; Criss et al., 2011) and recall (e.g. Wilson et al., 2020) in younger adults, free recall in older adults (Smith, 1975), and was replicated in cued recall for young adults in a pilot study for this experiment. However, hypotheses regarding the *phenomenon* of output interference have yet to be empirically validated against metamemory outcomes.

Of particular interest are both a) the specific question of whether metamemory accuracy during interference changes with age (e.g. Eakin & Hertzog, 2006, 2012a, 2012b) and b) the broader question of how the retrieval operations that underlie OI might be uncovered through learners' experiences during cued recall. This study explores these questions using a variation of the canonical Recall-Judge-Recognize procedure (Hart, 1965): Individuals were asked to study either completely unrelated cue and target word pairs (i.e. control items) or word pairs in which the cue and target pairs are unrelated to each other, but cue words are related to each other via taxonomic categories (i.e. "interference" items). After a short distractor task, individuals were asked to recall the target words that were co-presented with the cues during study. During each cued-recall trial, individuals rated a) whether they "Remember" or "Know" the target word (i.e. R/K/N judgments) and b) how confident they are that they will be able to recognize the correct target if they saw it on the screen (i.e. FOK judgments). In the final section of the experiment, individuals were asked to recognize the correct target word when it was co-presented with the cue and 3 distractors. After each 4AFC recognition trial, participants rated their confidence that they chose the correct target (i.e. RCJ).

Examining metamemory in the context of aging and OI will help to resolve a number of different theoretical issues. First, this study will help validate mechanistic accounts of OI by connecting declines in memory performance over trials to subjective metamemorial experiences. Anderson and colleagues' (M. C. Anderson et al., 2000; M. C. Anderson et al., 1994; M. C. Anderson & Neely, 1996) hypothesis of semantic competition driving OI effects is compatible with prior metamemory research in retroactive interference (Eakin, 2005; Maki, 1999), proactive interference (Diaz & Benjamin, 2011; Wahlheim, 2011), and implicit interference (Eakin & Hertzog, 2006, 2012a; Schreiber, 1998; Schreiber & Nelson, 1998); however, studies examining the role of inhibition in interference paradigms have yielded mixed results (e.g., evidence for inhibitory deficits; M. C. Anderson, 2003; Aslan & Bäuml, 2011; Hulbert et al., 2012; evidence against inhibitory deficits Aslan et al., 2007; Williams & Zacks, 2001). FOK and R/K/N judgments that learners produce in this experiment can be used to better understand mechanisms of interference in learners and whether the mechanisms that underlie OI are different between young and older adults. This information can also be connected to computational accounts of OI (e.g. Wilson et al., 2020) in order to better validate subjective (e.g. FOKs) and mechanistic (e.g. incrementing) accounts of interference. Second, this study will help explore the extent to which older and young adults show sensitivity to memory access during OI. If OI operates similarly to other interference states (e.g. Eakin & Hertzog, 2006, 2012a), then the availability of non-criterial recollection of information related to interference target items should remain intact, allowing for above-chance accuracy between FOKs made during recall and later recognition memory accuracy. However, the benefits of non-criterial recollection for older adult learners may be overshadowed by a reduced ability to inhibit the intrusion of competitors during retrieval (M. C. Anderson, 2003). If target representations are occluded by previously-retrieved items, however, then access to non-criterial information will not be possible, resulting in chance FOK resolution for interference items. Lastly, this study will continue a line of research examining whether memory awareness (via FOKs) during interference—a strong cause of forgetting over the lifespan (M. C. Anderson & Neely, 1996)—is impaired with age. OI presents an optimal environment to explore this line of inquiry because OI effects on cued-recall have been shown to be qualitatively similar between young and older adults (Smith, 1975). In contrast, the studies by Eakin and Hertzog (2006, 2012a) show qualitative differences in interference between young and older adults (e.g. interaction effects with Age), making age-related metamemory comparisons difficult.

While there are a number of different hypothesized mechanisms of interference (c.f. M. C. Anderson & Neely, 1996), the results of the proposed study will be validated against two accounts:

1. *Competition Hypothesis*: Output interference is hypothesized to be a function of competition between co-activated items. In order to over come this competition, inhibition of co-activated competitors in early trials carries over to later trials, reducing the likelihood of retrieval over serial position (M. C. Anderson, 2003; M. C. Anderson et al., 1994; M. C. Anderson & Levy, 2007; M. C. Anderson & Neely, 1996). Uner this account, older adults should display a decreased ability to inhibit irrelevant competitors in OI (Hasher et al., 1991; Hasher & Zacks, 1988; Hasher et al., 1999).

2. *Retrieval-suppression Hypothesis*: Output interference is hypothesized to be the function of related, but previously-activated items occluding conscious retrieval of a target item later in a retrieval sequence. Put differently, the previously-items should "overshadow" the correct target rather than compete with it in a spreading-activation system. Overcoming suppression of the target is not thought to be dependent upon actively inhibiting competitors (Bäuml, 1998; Williams & Zacks, 2001), so OI effects should be qualitatively similar between age groups (Aslan et al., 2007).

These two accounts are not necessarily in opposition to each other (c.f. Williams & Zacks, 2001), but in the context of this study, they predict different results.

Cued Recall. Given the similar retention interval for both age groups in this study, as well as expected age-related differences in recall performance (c.f. Kausler, 1994), I expect younger adults to have greater recall accuracy on average compared to older adults, regardless of whether cue words are related or not. In line with the retrieval-suppression hypothesis, I also expect that all participants will see a 0.5%-1% decrease in average recall for each successive cued recall trial, but only when the cue words from the cue-target pairs are semantically-related. The output interference effect during recall should be the same for both young and older adults, similar to the findings of Smith (1975); however, if older adults' OI effects are qualitatively different from those of young adults (e.g. low performance and no decline over output position), then a competition account would be a more appropriate explanation of recall, suggesting that LTM in older adults suffers from inhibitory deficits at the time of retrieval.

Recollection and Familiarity. Interference due to cue-relatedness should result in significantly fewer reports of recollective memorial experiences during recall (i.e. *R* judgments) compared to participants who study completely unrelated cues and target words. This will be particularly true for older adult participants, as age-related decreases in memory performance are often marked by decreases in recollection (c.f. Perfect & Dasgupta, 1997). Condition-level differences in *R* judgments are expected to be attributed to increased interference in the Related Cues condition, resulting in decreases in *R* judgments that track increasingly impaired cued recall performance. Participants in the Unrelated Cues condition, however, should have relatively stable *R* rates across recall trials.

A critical set of results concerns the patterns of self-reported familiarity (*K*) and "no-memory" (*N*) memory states across cued-recall trials. For participants in the Unrelated Cues condition, *K* and *N* judgment rates should remain stable over trials (similar to *R* judgment rates). For participants that experience OI, however, *K* and *N* judgments rates are expected to exhibit one of two patterns: 1) *K* judgments increase over cued recall trials, suggesting that participants continue to have trace access to target items; or 2) *N* judgments increase over cued recall trials, suggesting that OI states block the representation of target items during recall, particularly later in the recall sequence. The first set of *K* and *N* judgment patterns would be more consistent with a competition hypothesis while the second set would be more consistent with a retrieval-suppression account.

Feelings-of-knowing. A critical concern of this study is the extent to which interference affects metamemory judgment accuracy. Here, feelings-of-knowing (FOKs), which are made during cued-recall, will be used to measure the extent to which learners believe they have (or will have) access to a particular target item during recognition. In general, patterns of FOK magnitude (i.e. averaged FOKs) should closely match cued recall performance, such that mean FOKs stay stable over recall trials in the Unrelated Cues condition, but consistently decrease over trials in the Related Cues condition. This should be similar for both age groups, consistent with previous studies that found a large correspondence between FOKs and recall outcomes in both young and older adults, even under interference conditions (e.g. Eakin & Hertzog, 2012a).

Importantly, the degree to which FOK judgments correspond with later memory trials, i.e. FOK resolution, will help elucidate how OI affects metamemory. The overall and age-related results are expected to be dependent on the degree of criterial and non-criterial information learners will be able to access during OI:

- A retrieval-suppression account hypothesizes that access to a target item is occluded by related items that were retrieved earlier in a memory test. While this may not occlude the *entire* representation, this suppression would mask most intentional recollection of the target item such that FOKs for related cues and targets decrease in resolution (compared to later recognition memory performance) across cued-recall trials. Without the help of elaborative or distinctiveness encoding, these FOKs will be unable to be constructed with information via non-criterial recollection.

- A competition account hypothesizes that target and related competitor items are co-activated simultaneously, with greater activation for items that have been recently retrieved. While irrelevant items that are more highly-activated will be salient during retrieval, access to the correct target should *not* be blocked; instead, FOKs should reflect a feeling that the target item is "there" (activated), but that further context is needed to fully recollect the item. Here, FOKs should not significantly decrease in resolution across trials for items with related cues and targets.

The extant literature is unclear as to whether there will be age-related differences in FOK accuracy for an OI task (outside of mean-level differences in judgment magnitude). Implicit interference research suggests that older adults are able to give accurate item-level FOKs during competition at retrieval (implicating a competition account; Eakin & Hertzog, 2006, 2012a); however, it is also possible that the the proposed task will negatively bias FOKs for older adults that experience occlusion by decreasing confidence in their overall memory performance (c.f. Kelley & Sahakyan, 2003). Any such age-related differences will be reflected in the degree to which young and older adults differ in their FOK resolution estimates during OI.

While the proposed project is a novel concept without direct precedent in the extant literature, the results will be informative of the degree to which learners are aware of their own memory during a common interference task (OI) and will help aging researchers understand whether metaememory deficits in older adults can be partially attributed to qualitative differences in interference states.

CHAPTER 2
METHODS

2.1 Participant Characteristics

Eighty-five college-aged students (M_{age} = 20.16, SD_{age} = 3.95, N_{female} = 49) were recruited from the Georgia Institute of Technology's research subject pool and given credit for their participation. Of these young adult participants, 49 were randomly-assigned to the Related Cues condition while 36 were assigned to the Unrelated Cues condition. Additionally, 66 older adult participants ($Range_{age}$ = 55-73[1], M_{age} = 60.52, SD_{age} = 8.86, N_{female} = 41) were recruited via Amazon's Mechanical Turk (MTurk) platform and screened for medical issues that might infere with memory and decision-making in the experiment. Of these older adults, 35 were randomly-assigned to the Related Cues condition while 31 were assigned to the Unrelated Cues condition. Older adults were paid $10.00 upon completion of the study. All participants completed a short demographic questionnaire, along with two normed, general-intelligence tasks—pattern comparison and vocabulary—in order to assess individual differences in processing speed and verbal knowledge, respectively. All participants performed the task online using a PsychoPy (Peirce et al., 2019) program delivered through Qualtrics and MTurk.

The older adults recruited through MTurk were well-educated, with the vast majority having a college degree or greater (Years of Schooling: M = 15.00, SE = 0.31) and reporting proficiency with computers and technology. All older adults were native English speakers, with only one native English speaker who was born outside of the United States. All participants had good or corrected vision and did not self-report any disorder that could

[1] Amazon's Mechanical Turk allows researchers to target specific participants through "Premium Qualifications", where tasks become available to workers that meet a specific criteria. The highest age Qualification is "Age 55 or older" and cannot be amended. As such, the average age for older adults is lower than in most cross-sectional aging studies.

Age	Condition	N	Age M	Age SE	Shipley Vocabulary M	Shipley Vocabulary SE	Pattern Comparison M	Pattern Comparison SE
YA	Related	49	20.24	*0.57*	31.65	*0.50*	35.69	*1.24*
	Unrelated	36	19.73	*0.25*	29.68	*0.71*	36.00	*1.18*
OA	Related	35	61.13	*1.20*	34.13	*1.12*	28.15	*1.15*
	Unrelated	31	60.23	*1.55*	35.03	*1.14*	25.29	*1.16*

Table 2.1: Means and standard errors of participant characteristics.

affect cognitive performance (e.g., Alzheimer's disease, stroke, or mild cognitive impairment).

With respect to normative intelligence, participants did not differ in average task performance between the Related and Unrelated Cues conditions; however, the two age groups did differ in both the Shipley Vocabulary and Pattern Comparison tasks. For the Shipley Vocabulary task, where participants are asked to identify the closest synonyms for 40 words of increasing difficulty, a 2 (Age Group: Young Adults vs. Older Adults) by 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA only yielded a significant main effect of Age Group, F(1,147) = 18.39, $p < 0.01$, η_p^2 = 0.11, where older adults, M = 34.55, SE = 0.86, outperformed young adults, M = 30.82, SE = 0.38. Conversely, when examining Pattern Comparison task performance, where participants are asked to judge the similarity of 2 sets of 30 line drawings as quickly as possible (i.e. 60 total trials), a 2 (Age Group: Young Adults vs. Older Adults) by 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA with only a significant main effect of Age Group, F(1,147) = 61.19, $p < 0.01$, η_p^2 = 0.29, shows that young adults, M = 35.82, SE = 0.74, significantly outperform older adults, M = 26.80, SE = 0.90.

2.2 Materials

The study is a 2 x 2 quasi-experimental design. Age (Young vs. Older Adults) and Condition (Related vs. Unrelated Cues) are between-subjects variables, where individuals in the Related Cues condition will study lists of words with categorically-related cues, but unrelated targets, and those in the Unrelated Cues condition will study lists of words with both unrelated cues and unrelated cue-target pairs.

Related cue words were constructed from taxonomic category norms collected by the authors (Hertzog et al., 2019) based on previous category norms (Battig & Montague, 1969; Van Overschelde et al., 2004). These word lists consist of select exemplars from 20 categories, where the most- and least-typical exemplars are excluded. Categories with members that are difficult to differentiate or that are region-specific (e.g., *Cities*) have been excluded. Target words in the Related Cues condition, as well as both cues and targets in the Unrelated Cues condition, were sampled from a separate list of exemplars randomly-selected from the University of South Florida's Free Association Norms (D. L. Nelson et al., 2004).

Memory and metamemory performance can be significantly influenced by semantic properties that exist between studied words (Eakin & Hertzog, 2006, 2012a, 2012b; Schreiber, 1998; Schreiber & Nelson, 1998); therefore, extra procedures were necessary to eliminate potential contamination:

1. All potential category exemplars that overlap with the selected category exemplars were excluded from the unrelated, randomly-selected exemplars.

2. Using ListCheckerPro 1.2 (Eakin, 2010), I compared normative semantic associations between potential cues (i.e. category exemplars) and the randomly-selected words from the USF Free Association Norms. All words with forwards- or backwards-associations with any category exemplar greater than or equal to 0.05 (c.f. D. L. Nelson et al., 2004) were excluded.

3. I also compared all potential cue/target words selected from the USF Norms to each other using ListCheckerPro. If two words had forwards- or backwards-associations greater than or equal to 0.05, one of the two words was randomly chosen to be removed from the list. This technique helps eliminate unwanted competition between words in any given stimulus set during the task.

4. Next, the remaining words were again compared against each other in ListChecker-Pro. Here, word pairs were flagged if they shared a strong relationship with a third word ("mediator"). If a mediating relationship greater than or equal to 0.1 existed between a category exemplar and a randomly-selected word from the USF Norms, then the latter word was removed. If a mediating relationship existed between two randomly-selected words from the USF Norms, then one word was randomly selected to be removed.

5. Finally, any word pairs with a mediating relationship greater than 0.0, but less than 0.1, were flagged and added to an exclusion list, such that the experimental program will not pair them together. For example, the "Type of Occupation“ exemplar *DENTIST* and randomly-selected word *POST* were flagged because of their shared relationship with *OFFICE*, which is not included in the potential word list. Excluding such items helps control for indirect associative effects on memory retrieval (D. L. Nelson & Zhang, 2000).

For each participant, 40 cue-target pairs were randomly selected as stimuli. For the Related Cues condition, 10 exemplars from each of 4 categories served as cues for paired associates and were coupled with a randomly-selected word from the USF Free Association Norms (using the exclusion criteria listed above). Thus, for these participants, only cue words, but not the cue-target pairs themselves, were related to each other. For the Unrelated Cues condition, 80 randomly-selected exemplars from the list of unrelated words were used to construct the 40 cue-target pairs, with the provision that no exemplars are repeated. A Python script within the experimental program randomly selected the categories, category exemplars, and unrelated words for each participant.

2.3 Procedure

At study, participants were instructed to study each cue-target pair for a memory test later on. The instructions did not disclose the nature of the stimuli or the experiment. All participants had 4 s to study each word pair. All 40 items were presented in a random order for all participants in two blocks[2], equalling 80 study trials. After study, all participants engaged in a 5 minute distractor task before moving on to the test portion of the experiment.

The second phase of the experiment consisted of two procedures: recall and recognition (Hart, 1965). During the cued recall phase, participants were shown a cue word from study on the screen and asked to try and recall the word that completed the cue-target pair by typing their response. After each recall response, participants were asked to give a feeling-of-knowing judgment regarding their confidence that they will be able to recognize the correct target word in a recognition test. Using a mouse, participants reported FOKs using a continuous scale (i.e. "slider") between 0 ("No confidence in recognizing the correct target") to 100 ("Absolute confidence in recognizing the correct target"). Participants were also asked to report if they felt that they "Remember" a given item (i.e. recollect specific details about the target), "Know" what the target is (i.e. a feeling of familiarity without specific recollection of details), or have "No Memory" (referred to as R/K/N judgments). The experimental program queried for FOK and R/K/N judgments immediately after a recall response, or if the recall trial exceeded the maximum time (10 s).

Participants in the Related Cues condition experienced 4 back-to-back blocks of cued recall: One for each set of 10 cue-target pairs whose cues share a semantic relationship, i.e. normative taxonomic category. Because all items were randomly-chosen for participants in the Unrelated Cues condition, the order of presentation during recall for these individuals was randomized using the experimental software.

[2]Pilot data collected before the main data collection suggests that only one exposure to the word pairs results in low cued-recall performance. Thus, participants studied each word pair twice in order to bring recall performance off of floor. M. C. Anderson et al. (2000) report that study repetition does not influence retrieval-induced forgetting.

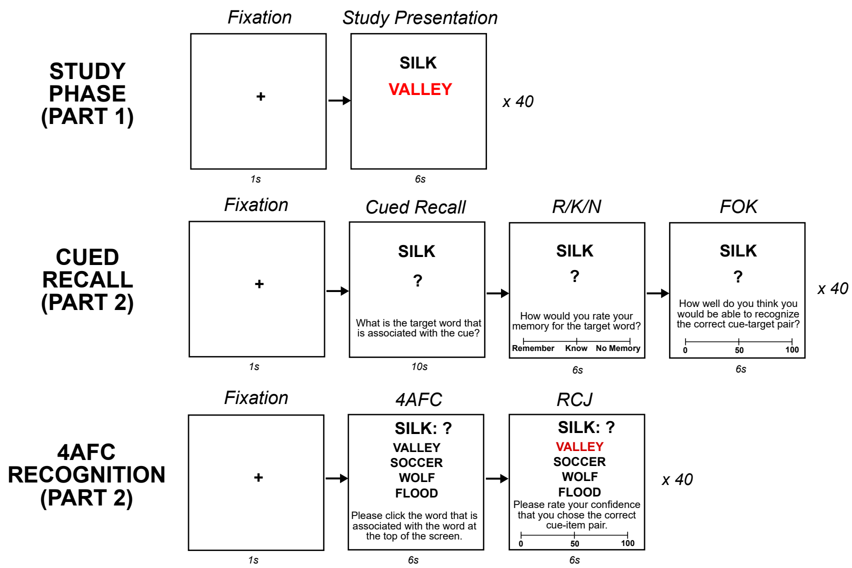

Figure 2.1: A visual example of the experimental procedure.

After recall, participants engaged in a four-alternative forced-choice (4AFC) task, where all items were presented in a random order. During each 4AFC trial, a previously-studied cue word along with 4 potential targets were shown on the screen: the correct target word, a previously-studied word from a separate pair, and two new, unseen words. Participants were asked to click on the word that was presented with the cue word during study. The position of each of the 4 alternatives on the screen was randomly selected for each trial. After each recognition trial, participants were asked to give a retrospective confidence judgment (RCJ) indicating their confidence that they chose the correct target. Judgments were given on a continuous scale, where 0 indicates no confidence in recognizing the correct target and 100 indicates absolute confidence in recognizing the correct target. Like FOKs, participants gave their RCJs by clicking on a labeled sliding scale using a mouse.

CHAPTER 3

RESULTS

3.1 Statistical Methods

All statistical analyses were performed using the R statistical programming language (R Core Team, 2020).[1] Several packages within this environment were employed to perform these analyses. I used the `BayesFactor` package (Morey & Rouder, 2018) to test the evidence for (BF_{01}) and against (BF_{10}) the a priori null hypotheses using Bayes Factor estimates (c.f. Brydges & Bielak, 2020) where appropriate. The specific formulations of the null and alternative hypotheses for each analysis are listed in the appropriate sections of this chapter. In line with Raftery's (1995) recommendations, I interpret these Bayes Factors as:

- **BF = 1** - No evidence.
- **1** $<$ **BF** $\leq$ **3** - Weak evidence.
- **3** $<$ **BF** $\leq$ **10** - Positive evidence.
- **10** $<$ **BF** $\leq$ **150** - Strong evidence.
- **BF** $>$ **150** - Extreme evidence.

Additionally, I used the `lme4` package (Bates et al., 2015) to run multi-level models, specifically the `lmer()` command for linear mixed models and `glmer()` with a logit link for binomial mixed models. Unless otherwise specified, these models used the default optimization settings. For linear mixed models, I report estimates of model fit for the full model with respect to a) only the fixed effects (γ) and b) both the fixed (γ) and random (τ)

[1]The materials and code used to perform these analyses are available through the Open Science Framework: https://osf.io/cbpfj/.

effects using estimates of their variance. In a linear multilevel model, for example, model fit can be expressed as:

$$R^2_{Fixed} = \frac{\gamma^2}{\gamma^2 + \tau^2 + \sigma^2}$$

$$R^2_{Overall} = \frac{\gamma^2 + \tau^2}{\gamma^2 + \tau^2 + \sigma^2}$$

where σ^2 is the residual variance in the model. Thus, R^2 can be interpreted similarly to fit statistics from ordinary linear regression, i.e. percent of estimated variance in a model. For three-level models, this R^2 formulation can be extended to estimate variance accounted for by the second (γ) and third (ϕ) levels.

For logistic multilevel models, where parameter estimation is not based on ordinary-least-squares (OLS) methods (Hosmer et al., 2013), I instead use the pseudo-R^2 metric described by Snijders and Bosker (2011), where model fit is defined in terms of explained (σ^2_F) and unexplained ($\tau^2 + \sigma^2_R$) variance, i.e.:

$$R^2_{SB} = \frac{\sigma^2_F}{\sigma^2_F + \tau^2_0 + \sigma^2_R}$$

where τ^2_0 is the intercept variance and σ^2_R is level-one residual variance. For logistic mixed models, the level one residual variance is fixed to $\pi^2/3$ (Snijders & Bosker, 2011). Note that pseudo-R^2 values cannot be directly interpreted as explained variance in a model and are typically considerably lower than OLS R^2 values obtained from linear multilevel models (Hosmer et al., 2013).

Young adults in the Related Cues condition were set as the reference group in each GLM unless otherwise specified.

3.2 Cued Recall

The central thesis of this project is predicated on the existence of the output interference effect in both young and older adults; thus, an analysis of cued recall performance

		% Recall		% Recog.		Mean FOK		Mean RCJ	
Age	Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
YA	Related	0.36	*0.04*	0.67	*0.03*	57.12	*2.88*	68.29	*2.33*
	Unrelated	0.38	*0.04*	0.65	*0.04*	56.98	*3.48*	66.60	*3.07*
OA	Related	0.54	*0.05*	0.71	*0.03*	67.16	*3.79*	71.06	*2.97*
	Unrelated	0.36	*0.04*	0.65	*0.03*	53.16	*4.24*	65.61	*3.37*

Table 3.1: Memory performance and judgment means by age group and experimental condition.

across participants is crucial. Based on wide-spread evidence of age-related declines in LTM (Kausler, 1994), I hypothesized that young adults would have higher overall cued-recall accuracy than older adults, but that no other main effects or interactions with respect to age groups and experimental conditions would be significant. Additionally, recall is expected to decrease across successive recall trials for both young and older adults, but only when cues from word pairs are semantically related, consistent with previous output interference research. As a result of consistent decreases in cued recall performance over trials, error rates are expected to increase, and a particular question of interest is whether omission error rates increase, indicating that retrieval is occluded (i.e. a *Retrieval-suppression Hypothesis*), or if commission error rates increase, indicating that retrieval is affected by semantic interference (i.e. a *Competition Hypothesis*).

3.2.1 Overall Means

I conducted a 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA on overall cued-recall accuracy (Figure 3.1). There was a significant Age Group by Condition interaction, $F(1,147) = 5.24$, $p = 0.02$, $\eta_p^2 = 0.03$. Follow-up analyses using Tukey's HSD indicates that older adults in the Related Cues condition had significantly higher recall accuracy, $M = 0.54$, $SE = 0.05$, than did older adults in the Unrelated Cues condition, $M = 0.36$, $SE = 0.04$, $p = 0.03$. Older adults in the Related Cues condition also significantly outperformed young adults in the same condition, $M = 0.36$, $SE = 0.04$, $p = 0.02$, but not young adults in the Unrelated Cues

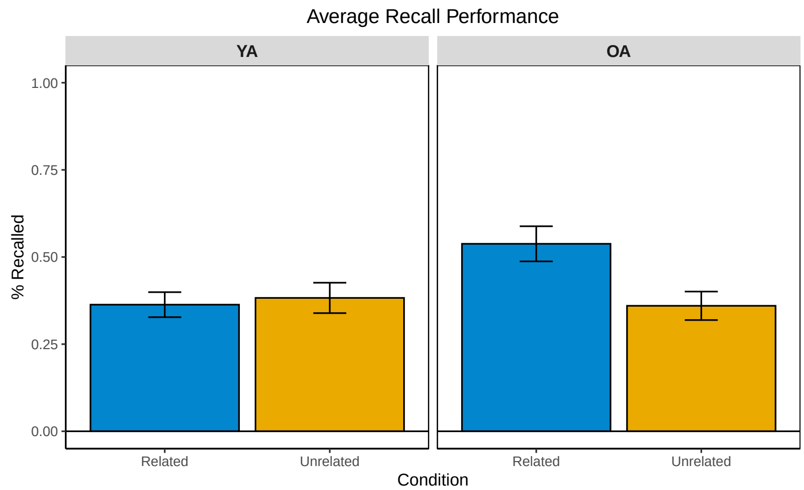

Figure 3.1: Average cued recall performance by age group and cue-relatedness.

condition, *M* = 0.38, *SE* = 0.04, *p* = 0.06. No other pairwise comparisons were significant.

Additionally, the main effect of Age Group was significant, *F*(1,147) = 9.14, $p < 0.05$, η_p^2 = 0.03, indicating that, on average, older adults had higher cued recall accuracy, *M* = 0.45, *SE* = 0.03, than did young adults, *M* = 0.37, *SE* = 0.03. The main effect of Condition was not significant, *F*(1,147) = 0.11, *p* = 0.74.

I calculated the degree of evidence in favor of the hypotheses of a main effect of Condition (H_1) and an interaction between Age Group and Condition (H_2) against the null hypothesis that a main effect of Age Group best fits the data (H_0). The null hypothesis model provides a better fit to the data than model with a main effect of Condition (BF_{01} = 1.96), but not the interaction model, which is 1.93 times more likely than the null model (BF_{03} = 0.51). The Bayes Factors in these analyses only point to anecdotal evidence in favor to the null or alternative hypotheses stated above.

Binomial logistic regression. I additionally ran a logistic regression on the cued recall outcome data ("recalled" = 1, "not recalled" = 0) to verify the results of the ANOVA reported above. A model with Condition (Unrelated Cues vs. Related Cues) and Age Group (Young vs. Older Adults) accounted for a significant amount of information compared to the null model, $\chi^2(3)$ = 128.02, $p < 0.01$. Only the interaction between Age Group and

Condition was significant, resulting in a significant parameter estimate for older adults in the Unrelated Cues condition, Exp(B) = 0.44, z = -7.55, $p < 0.01$. Pairwise z-ratio tests indicate that the interaction is carried by Older Adults in the Related Cues condition, Exp(B) = 0.54, SE = 0.01, who have a significantly greater chance at correctly recalling target words compared to Young Adults in the Related Cues condition, Exp(B) = 0.36, SE = 0.01, z = -10.01, $p < 0.01$, Young Adults in the Unrelated Cues condition, Exp(B) = 0.38, SE = 0.01, z = 8.26, $p < 0.01$, and Older Adults in the Unrelated Cues condition, Exp(B) = 0.36, SE = 0.01, z = 9.11, $p < 0.01$. No other effects were significant.

3.2.2 Recall Accuracy Across Trials

In this study, evidence in favor of an output interference effect would present as declines in average recall performance across trials within each block of trials. As a preliminary test to this hypothesis, I averaged cued recall performance for each participant in the Related Cues condition across the 10 trials in each of the 4 blocks of cued recall that they experienced during test (Figure 3.3). For participants in the Unrelated Cues condition, who did not experience blocked recall trials, I divided the 40 cued recall trials into 4 "blocks" of 10 successive trials each and averaged across trials within these pseudo-blocks. This allows for a direct comparison between groups, where one is expected to show linear[2] decreases in cued recall accuracy across trials (Related Cues condition), while the other is not (Unrelated Cues condition).

The means were submitted to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Unrelated Cues vs. Related Cues) x 10 (Trial Within Recall Cycle: 1 - 10) mixed-effects ANOVA with repeated measures on the last factor. Mauchley's Test is significant, W =

[2]There is no evidence to suggest that the output interference effect is strictly linear; in fact, recent reports of output interference in cued recall are purposefully agnostic to the form of the function and simply test for decreases across trials using Bayes Factor analyses (Wilson et al., 2020). A linear form was chosen for these analyses because the output interference effect should yield a significant negative linear trend in a typical analysis using the GLM, regardless of the highest degree. Determining the true form of the output interference effect in cued recall is beyond the scope of this project; however, additional analyses on cued recall rates across trials using higher-order functions yielded non-significant results (Appendix 2.1).

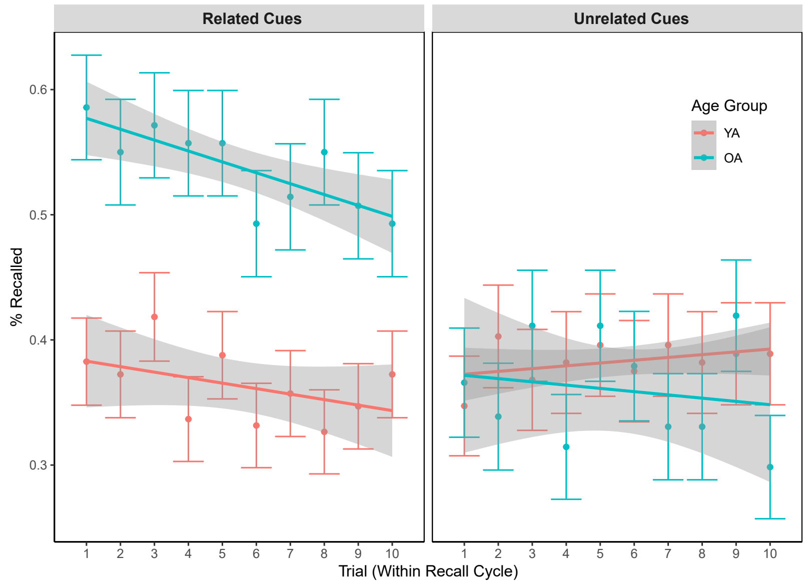

Figure 3.2: Average cued recall performance by age group and cue-relatedness averaged across blocked trials.

0.647, p = 0.03, indicating that the analysis violates the assumption of sphericity. Only the Age Group by Condition interaction is significant, F(1,147) = 5.25, p = 0.02, η_p^2 = 0.02. A Bonferroni-corrected paired-samples test on the 4 levels of this interaction indicates that the effect is carried by higher cued recall accuracy for older adults in the Related Cues condition, M = 0.54, SE = 0.02, compared to older adults in the Unrelated Cues condition, M = 0.36, SE = 0.02, $p < 0.01$, and young adults in both the Related Cues, M = 0.36, SE = 0.01, $p < 0.01$, and Unrelated Cues, M = 0.38, SE = 0.02, $p < 0.01$, conditions. No other paired comparisons were significant, nor were any other main effects, although the main effects of Condition, F(1,147) = 3.39, p = 0.07, η_p^2 = 0.01, and Age Group, F(1,147) = 3.11, p = 0.08, η_p^2 = 0.01, approached significance.

Despite the non-significance of the 3-way interaction in this model, F(0.92, 8.25) = 0.83, p = 0.58, I decided to test the simple slopes for each of the age groups and conditions across recall trials, averaged across the 4 recall blocks. The simple slope for older adults in the Related Cues condition was negative and significantly different than zero, B = -0.009, t = -2.66, p = 0.01, indicating a significant decrease in average recall performance over trials.

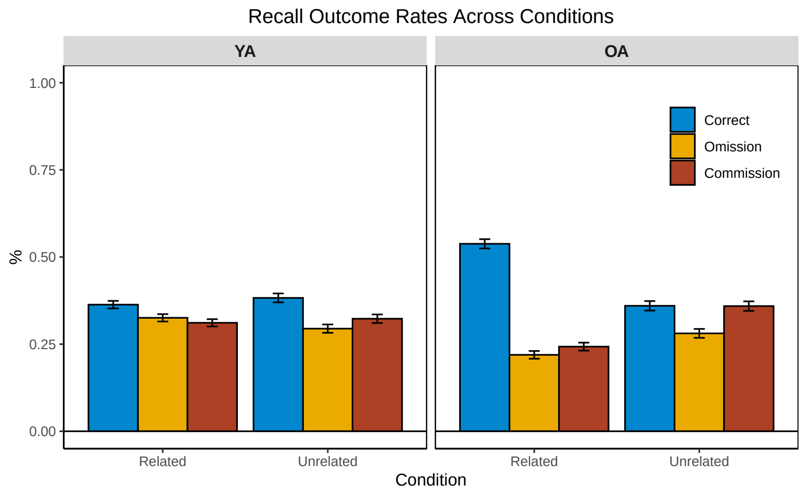

Figure 3.3: Correct recall, omission, and commission rates across cue-relatedness and age conditions.

The same simple slopes for both older adults, B = -0.003, t = 0.43, and young adults, B = 0.002, t = 0.68, p = 0.50, in the Unrelated Cues condition were not significantly different than zero, nor was the simple slope for young adults in the Related Cues condition, B = -0.004, t = -1.34, p = 0.19.

When the data are averaged across age groups, however, the simple slope for all participants in the Related Cues condition is negative and significantly different than zero, B = -0.007, t = -2.83, $p < 0.01$, whereas the same simple slope for participants in the Unrelated Cues condition was not different than zero, B = -0.0002, t = -0.08, p = 0.94.

3.2.3 Recall Outcomes

An equally important question in this study is the pattern of errors that participants exhibit during output interference. To examine this, I coded cued recall responses in one of 3 ways: "Correct" for correctly-recalled target words; "Omission" for trials in which the recall response was blank or included exclusionary words (e.g. "pass", "I don't know", "next"); and "Commission" for trials in which a valid response was given, but the response is incorrect. Figure 3.3 shows the average occurrence rates for each of the 3 outcomes across experimental conditions.

Means Across Conditions. I first submitted the average recall outcome occurrence rates to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) x 3 (Outcome: Correct Recall vs. Omission vs. Commission) mixed-effects ANOVA, with repeated measures on the last factor. Only the main effect of Outcome was significant, F(2,294) = 6.38, p = 0.002, η^2 = 0.04. Bonferroni-corrected paired-samples t-tests on the Outcome factor indicate that the rate of correct recall, M = 0.41, SE = 0.04, is significantly greater than that of errors of omission, M = 0.28, SE = 0.28, p = 0.04. Correct and commission error rates, M = 0.31, SE = 0.05, did not significantly differ, p = 0.14, nor did omission and commission error rates, p = 1.0. Mauchley's Test indicates that the assumption of sphericity holds in this analysis, W = 0.99, p = 0.64.

Multilevel Model

The multilevel modeling framework is a prime candidate for simultaneously accounting for group-, participant-, and cycle-level variation in recall outcomes (i.e. correct recall, error of omission, and error of commission). However, having three levels to an outcome variable presents a unique issue: The number of categories is just small enough that the distribution of the residuals cannot be approximated by a normal distribution (Snijders & Bosker, 2011); thus, a linear hierarchical model is not appropriate for these data. For categorical dependent variables with 3 to 5 levels, Snijders and Bosker (2011) recommend dichotomizing the dependent variable. This simplifies both the computation and interpretation of the resulting parameters (although at the cost of some information full ordinal multilevel models might otherwise have; 2011). Here, I describe 3 separate binomial models using a logit link function: Recalled vs. {Omission, Commission}, Omission vs. {Recalled, Commission}, and Commission vs. {Recalled, Omission}. The parameters of these models can be interpreted as the likelihood of a given recall outcome (e.g. correct recall) versus all other outcomes (e.g. errors of omission and commission). These models are analyzed across 3 levels: The level of the individual trial, the level of the cycle within the cued recall test, and the level of the individual participant.

Random Components. Before specifying the full models, I used a single-term dropout technique to determine whether the random effects in these models (i.e. random slopes for recall cycles and individual participants) account for a significant amount of information. A significant chi-squared test on the difference in deviance values between the model with the single-term deletion and a model with both random effects indicates that a random effect accounts for a significant amount of information in a given model. The exclusion of the random intercept of recall cycle was significant for the correct recall, $\chi(1)$ = 10.63, $p <$ 0.01, and commission models, $\chi(1)$ = 5.13, p = 0.02, but only trending towards significance for the omission model, $\chi(1)$ = 3.76, p = 0.052. The exclusion of the random intercept of participant was significant for all models: correct recall, $\chi(1)$ = 1453.50, $p <$ 0.01; omission, $\chi(1)$ = 1796.91, $p <$ 0.01; commission, $\chi(1)$ = 1757.64, $p <$ 0.01. Based on these results, I included the random terms for both recall cycle and participant in the multilevel models.

	Correct (1)		Omission (2)		Commission (3)	
	df	*F*	*df*	*F*	*df*	*F*
Age Group	1, 187.1	**5.53***	1, 166.2	**4.77***	1, 189.6	0.42
Condition	1, 187.1	**5.92***	1, 166.2	1.34	1, 189.6	2.79+
Cycle	3, 6007	0.09	3, 6007	1.20	3, 6007	0.07
Trial (Within Cycle)	1, 6007	3.26+	1, 6007	**10.68***	1, 6007	1.17
Age x Condition	1, 187.1	**4.06***	1, 166.2	1.19	1, 189.6	1.34
Age x Cycle	3, 6007	2.08	3, 6007	2.57	3, 6007	1.13
Age x Trial	1, 6007	1.40	1, 6007	2.57	1, 6007	<0.01
Condition x Cycle	3, 6007	**2.75***	3, 6007	**3.02***	3, 6007	0.18
Condition x Trial	1, 6007	2.93+	1, 6007	2.05	1, 6007	0.36
Cycle x Trial	3, 6007	1.90	3, 6007	0.66	3, 6007	2.51+
Age x Condition x Cycle	3, 6007	1.52	3, 6007	**2.73***	3, 6007	0.40
Age x Condition x Trial	1, 6007	<0.01	1, 6007	0.28	1, 6007	0.24
Age x Cycle x Trial	3, 6007	0.76	3, 6007	0.35	3, 6007	0.38
Condition x Cycle x Trial	3, 6007	1.62	3, 6007	1.06	3, 6007	0.17
Age x Condition x Cycle x Trial	3, 6007	2.00	3, 6007	**3.91***	3, 6007	0.51
Note:						$^{+}$p<0.1; *p<0.05

Table 3.2: *F*-table of the multilevel models predicting correct recall (Model 1), errors of omission (Model 2), and errors of Commission (Model 3).

Table 3.3: Parameter estimates for the multilevel models predicting correct recall (Model 1), errors of omission (Model 2), and errors of commission (Model 3). Results are reported in log-odds.

	Dependent variable:		
	Correct	Omission	Commission
	(1)	(2)	(3)
Intercept	0.006	−2.236*	−0.970*
	(0.276)	(0.374)	(0.311)
Age (OA)	0.282	−0.036	−0.459
	(0.433)	(0.587)	(0.502)
Condition (Unrelated)	−0.805+	0.864	0.267
	(0.428)	(0.561)	(0.483)
Cycle 2	**−1.151***	**1.424***	0.075
	(0.285)	(0.328)	(0.299)
Cycle 3	**−0.633***	**1.217***	−0.314
	(0.279)	(0.329)	(0.303)
Cycle 4	**−0.692***	**0.932***	0.031
	(0.282)	(0.329)	(0.297)
Trial (Within Cycle)	**−0.115***	**0.161***	−0.006
	(0.037)	(0.044)	(0.040)
Age (OA) x Condition (Unrelated)	−0.173	−0.562	0.747
	(0.646)	(0.858)	(0.735)
Age (OA) x Cycle 2	**1.329***	**−1.543***	−0.204
	(0.442)	(0.533)	(0.508)
Age (OA) x Cycle 3	0.757+	**−1.215***	0.139
	(0.439)	(0.532)	(0.508)
Age (OA) x Cycle 4	**1.149***	**−1.510***	−0.180
	(0.442)	(0.546)	(0.502)
Condition (Unrelated) x Cycle 2	**1.347***	**−1.812***	0.073
	(0.435)	(0.495)	(0.469)
Condition (Unrelated) x Cycle 3	0.837+	**−0.971***	−0.219
	(0.431)	(0.487)	(0.482)
Condition (Unrelated) x Cycle 4	0.615	−0.564	−0.335
	(0.436)	(0.485)	(0.472)
Age (OA) x Trial	0.102+	−0.116+	−0.018
	(0.058)	(0.070)	(0.067)
Condition (Unrelated) x Trial	**0.170***	**−0.197***	−0.028
	(0.057)	(0.066)	(0.063)
Cycle 2 x Trial	**0.215***	**−0.222***	−0.053
	(0.053)	(0.060)	(0.057)
Cycle 3 x Trial	**0.107***	**−0.198***	0.049
	(0.053)	(0.060)	(0.056)
Cycle 4 x Trial	0.034	−0.097	0.030

	(0.054)	(0.060)	(0.056)
Age (OA) x Condition (Unrelated) x Cycle 2	-1.107^{+}	**2.043***	-0.564
	(0.649)	(0.772)	(0.731)
Age (OA) x Condition (Unrelated) x Cycle 3	-0.959	1.358^{+}	0.065
	(0.650)	(0.763)	(0.734)
Age (OA) x Condition (Unrelated) x Cycle 4	-1.267^{+}	**1.772***	0.174
	(0.656)	(0.770)	(0.726)
Age (OA) x Condition (Unrelated) x Trial	**−0.178***	**0.246***	-0.003
	(0.086)	(0.102)	(0.097)
Age (OA) x Cycle 2 x Trial	**−0.225***	**0.280***	0.001
	(0.083)	(0.098)	(0.097)
Age (OA) x Cycle 3 x Trial	-0.160^{+}	**0.240***	-0.015
	(0.083)	(0.098)	(0.095)
Age (OA) x Cycle 4 x Trial	-0.120	0.160	0.039
	(0.083)	(0.100)	(0.094)
Condition (Unrelated) x Cycle 2 x Trial	**−0.259***	**0.341***	-0.018
	(0.081)	(0.092)	(0.090)
Condition (Unrelated) x Cycle 3 x Trial	**−0.162***	**0.227***	-0.007
	(0.081)	(0.091)	(0.090)
Condition (Unrelated) x Cycle 4 x Trial	-0.108	0.131	0.029
	(0.082)	(0.090)	(0.088)
Age (OA) x Condition (Unrelated) x Cycle 2 x Trial	**0.280***	**−0.477***	0.125
	(0.122)	(0.143)	(0.139)
Age (OA) x Condition (Unrelated) x Cycle 3 x Trial	0.222^{+}	**−0.338***	0.026
	(0.122)	(0.142)	(0.138)
Age (OA) x Condition (Unrelated) x Cycle 4 x Trial	0.204^{+}	-0.275^{+}	-0.042
	(0.123)	(0.142)	(0.136)
σ^2 (Recall Cycle)	<0.01	<0.01	<0.01
τ^2 (Participants)	1.37	1.93	1.58
-2LL	−3,318.53	−2,672.84	−2,834.24
R^2_{SB} (Fixed)	0.04	0.03	0.02
R^2_{SB} (Overall)	0.39	0.55	0.44

Note: $^{+}$p<0.1; *p<0.05

Correct Recall vs. Other. In a logistic multilevel model comparing cued recall trials with correct recall vs. all others, I included 3 categorical variables—age (Young vs Older Adults), condition (Related Cues vs. Unrelated Cues), and recall cycle (1 vs. 2 vs. 3 vs. 4)—and one continuous variable—trial within recall cycle (1 - 10)—where all variables are crossed (Table 3.3). An analysis using Snjider and Bosker's (2011) pseudo-R^2 indicates that the overall model accounts for a good amount of information in the data, R^2_{SB} = 0.39, but that the fixed effects only account for a small amount of information in the data, R^2_{SB} = 0.04. The results of the overall (F-tests) and parameter-specific (z-tests) tests of the fixed effects are reported below.

The 4-way interaction between age group, condition, cycle, and trial is not significant, F(3,6007) = 2.00, p = 0.11, despite a significant parameter estimate. Recall by recall trial (within cycle) parameter estimates for older adults in the unrelated condition for cycles 3, Exp(B) = 1.23, z = 1.65, p = 0.10, and 4, Exp(B) = 1.25, p = 1.82, p = 0.07, approach significance while the estimate for cycle 2 is positive and significant, Exp(B) = 1.32, z = 2.30, p = 0.02. A follow-up simple slopes analysis (Figure 3.4) indicates that the effect is driven by significant decreases in correct recall for young adults in the Related Cues conditions during cycle 1, Exp(b) = 0.89, z = -3.06, $p < 0.01$, and cycle 4, Exp(b) = 0.92, z = -2.07, p = 0.04, but a significant increase in cycle 2, Exp(b) = 1.11, z = 2.66, p = 0.01. Similarly, older adults in the Related Cues condition showed a significant decrease in correct recall across trials, but only during cycle 4, Exp(b) = 0.90, z - -2.21, p = 0.03. No other simple slopes are significant.

A 3-way analysis involving experimental condition, recall cycle, and trials within cycles, F(3,6007) = 1.62, p = 0.18, yielded similar results: While declines in cued recall accuracy for participants in the Unrelated Cues condition were only marginally significant in the second recall cycle, Exp(B) = 0.85, z = -2.0, p = 0.05, this trend for participants in the Unrelated Cues Condition was significant during the first recall cycle, Exp(B) = 0.77, z = -3.18, $p < 0.01$. A simple slopes analysis on this interaction indicates that significant

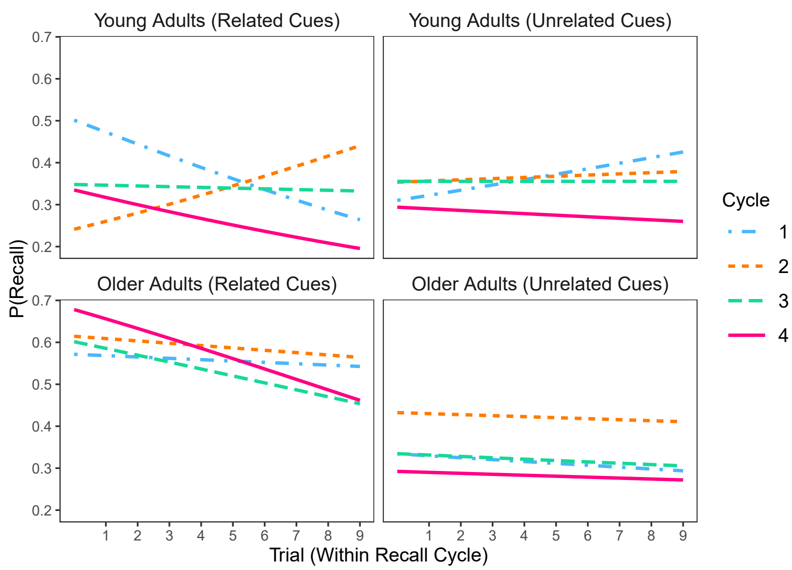

Figure 3.4: Simple slopes analysis estimating the probability of correct recall across recall trials within age group, condition, and recall cycle.

decreases in recall across trials only occur for participants in the Related Cues condition in cycle 1, Exp(b) = 0.93, z = -2.44, p = 0.01, and cycle 4, Exp(b) = 0.91, z = -3.02, $p < 0.01$. No other effects or simple slopes are significant.

The 3-way interaction between age group, cycle, and recall trial is not significant, F(3,6007) = 0.76, p = 0.94, despite significant parameter estimates. For older adults, the decrease in cued recall accuracy across recall trials is marginally significant during cycle 3, Exp(B) = 0.85, z = -1.94, p = 0.05, and fully significant during cycle 2, Exp(B) = 0.80, z = -2.72, $p < 0.01$. A simple slopes analysis indicates that young adults display increases in recall accuracy across trials during cycle 2, Exp(b) = 1.06, z = 2.16, p = 0.03, but decreases in cycle 4 approaching significance, Exp(b) = 0.95, z = -1.82, p = 0.07. For older adults, the recall accuracy by trial slope only approaches significance during the last cycle, Exp(b) = 0.94, z = -1.85, p = 0.06.

Similarly, the 3-way interaction between age group, condition, and trial is not significant, F(1,6007) < 0.01, p = 0.97. The recall accuracy by trial parameter estimate for older adults in the Unrelated Condition is significant, however, Exp(B) = 0.84, z = -2.06, p = 0.04, indicating that changes in recall performance across trials differ across age groups

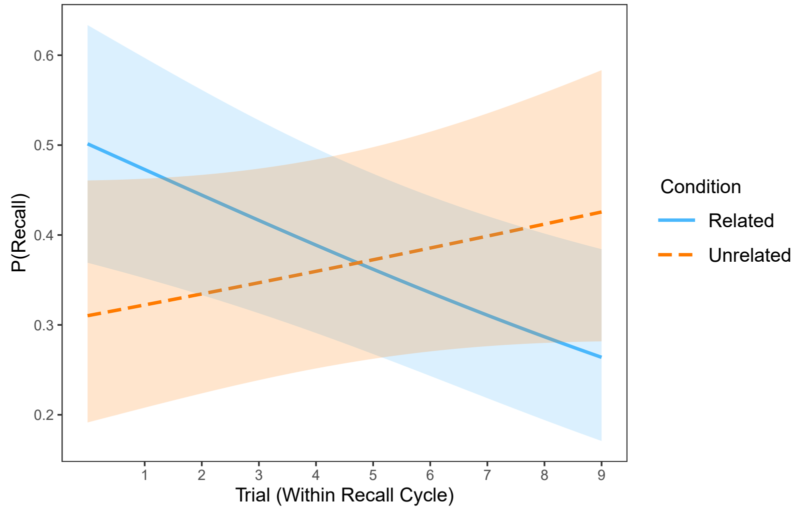

Figure 3.5: Simple slopes analysis estimating the probability of correct recall across recall trials for Related and Unrelated Cues conditions.

and experimental conditions. A follow-up analysis indicates that significant decreases in cued recall performance occur only for young adults in the Related Cues condition, Exp(b) = 0.90, z = -3.06, $p < 0.01$; no other simple slopes are significantly different than zero.

The interaction between cycle and trial within cycle is also not significant, F(3,6007) = 1.90, p = 0.13, despite containing significant parameter estimates. The significant parameter estimates for recall trial in cycle 3, Exp(B) = 1.11, z = 2.02, p = 0.04, and cycle 2, Exp(B) = 1.24, z = 4.04, $p < 0.01$, indicate that the recall by trial slopes change across recall blocks compared to cycle 1; however, a simple slopes analysis indicates that the recall by trial slope is significantly different than zero only for items in cycle 4, Exp(b) = 0.94, z = -2.58, p = 0.01.

The condition by trial interaction, F(1,6007) = 2.93, p = 0.09, which approaches significance, provides some evidence in favor of a general output interference effect, particularly with the significant recall by trial parameter estimate for participants in the Unrelated Cues condition, Exp(B) = 1.19, z = 2.97, $p < 0.01$. A simple slopes analysis on this interaction (Figure 3.5) indicates that while the trial by recall slope is not significantly different than zero for participants in the Unrelated Cues condition, Exp(b) = 1.02, z = 0.69, p = 0.49, it is for this in the Related Cues condition, Exp(b) = 0.93, z = -2.44, $p < 0.01$, indicating that the probability of correct recall decreases by a factor of 1.07, or 1.8%, per successive trial.

Average recall performance for a given recall block also differs between age groups, although the overall age group by cycle interaction only approaches significance, $F(3)$ = 2.08, p = 0.10. Both the parameter estimates for older adults in cycle 4, Exp(B) = 3.16, z = 2.60, $p < 0.01$, and cycle 2, Exp(B) = 3.78, z = 2.60, p = 0.01, are significant, with the parameter estimate for older adults in the third block only approaching significance, Exp(B) = 2.13, z = 1.73, p = 0.08. This age by cycle interaction is most likely due to older adults' significantly higher recall accuracy in the Related Cues condition.

As expected, the age group by condition interaction is significant, $F(1,187.1)$ = 4.06, p = 0.04. This reflects the higher probability of correct recall for older adults in the Related Cues condition, Exp(b) = 0.56, SE = 0.06, compared to older adults in the Unrelated Cues condition, Exp(b) = 0.50, SE = 0.06, z = 2.70, p = 0.04, and to young adults in both the Related Cues, Exp(b) = 0.48, SE = 0.04, z = 3.10, p = 0.01, and Unrelated Cues, Exp(b) = 0.51, SE = 0.05, z = 2.70, p = 0.04.

Across all cycles, recall performance tends to decrease across trials within blocks. This is evident through the overall main effect of trial, $F(1,6007)$, p = 0.07, which approaches significance, and its significant parameter estimate, Exp(B) = 0.89, z = -3.06, $p <$ = 0.01, which indicates that the likelihood of recall decreases by a factor of approximately 1.12, or 2.9%, per trial. Further, recall performance varies across blocked trials, despite a non-significant overall effect, $F(3,6007)$ = 0.21, p = 0.89, particularly for cycle 2, Exp(B) = 0.32, z = -4.04, $p < 0.01$, cycle 3, Exp(B) = 0.53, z = -2.27, p = 0.02, and cycle 4, Exp(B) = 0.50, z = -2.45, p = 0.01, when compared to cycle 1. Lastly, both the main effects of age group, $F(1, 187.1)$ = 5.53, p = 0.02, and condition are significant, $F(1,187.1)$ = 5.92, p = 0.02, and reflect the fact that older adult participants, Exp(b) = 0.80, SE = 0.04, were more likely to correctly recall target words than young adults, Exp(b) = 0.50, SE = 0.03, z = 2.01, p = 0.04, and that participants in the Related Cues condition, Exp(b) = 0.79, SE = 0.04, were more likely to correctly recall a target word than participants in the Unrelated

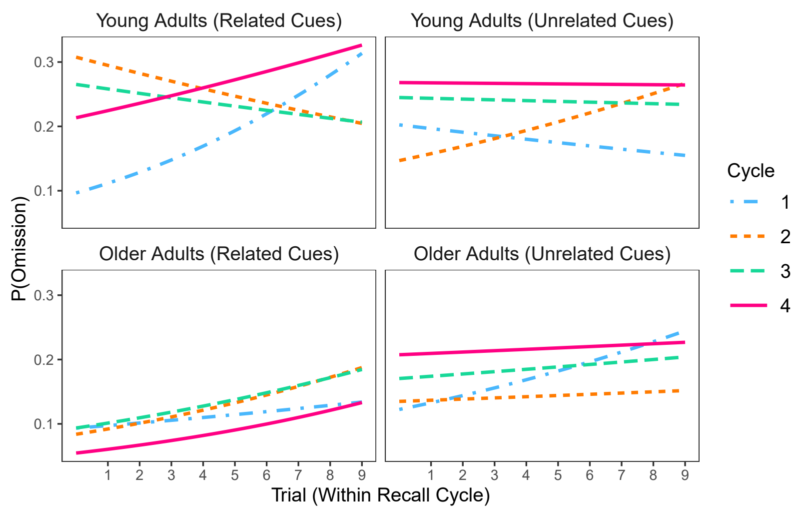

Figure 3.6: Simple slopes analysis estimating the probability of an error of omission across recall trials within age group, condition, and recall cycle.

Cues condition, Exp(b) = 0.51, SE = 0.04, z = 1.87, p = 0.06. No other parameter estimates or main effects are significant.

Errors of Omission vs. Other. I examined trial-level likelihoods of errors of omission using a similar logistic multilevel model framework where I included 3 categorical variables—age (Young vs Older Adults), condition (Related Cues vs. Unrelated Cues), and recall cycle (1 vs. 2 vs. 3 vs. 4)—and one continuous variable—trial within recall cycle (1 - 10)—where all variables are crossed (Table 3.3). An analysis using Snjider and Bosker's (2011) pseudo-R^2 indicates that the model accounts for a good amount of information, R^2_{SB} = 0.55, but that the fixed effects only account for a small amount of information in the data, R^2_{SB} = 0.03. Like before, the results of the overall (F-tests) and parameter-specific (z-tests) tests of the fixed effects are reported below.

The overall 4-way interaction between age group, condition, recall block (i.e. cycle), and trial within block, is significant, F(3,6007) = 3.91, p = 0.01, as are the individual parameter estimates for older adults in the Unrelated Cues condition during cycle 3, Exp(B) = 0.71, z = -2.39, p = 0.02, and cycle 4, Exp(B) = 0.62, z = -3.34, $p < 0.01$. A simple slopes analysis indicates that the interaction is carried by heterogeneity in the slopes for each cycle across the age groups and experimental conditions (Figure 3.6). Specifically, the simple slopes for omission errors across trials for older adults in the Related Cues condition during cycles 2, Exp(b) = 1.10, z = 1.88, p = 0.06, and 4, Exp(b) = 1.12, z = 1.89, p = 0.06, approaches significance, while the same simple slope for young adults in the Related Cues condition is significantly different than zero during cycle 1, Exp(b) = 1.17, z = 3.69, $p < 0.01$. These indicate that for participants in the Related Cues condition, the probability of omitting a response during recall increases by around 2% (older adults) to 4% (young adults) for each successive trial, depending on the block in the recall sequence. No other simple slopes are significant.

While the 3-way interaction between experimental condition, cycle, and trial within cycle is not significant, F(3,6007) = 1.06, p = 0.37, the parameter estimates for omission probabilities across trials for participants in the Unrelated Cues condition during cycles 2, Exp(B) = 1.41, z = 2.85, $p < 0.01$, and 3, Exp(B) = 1.25, z = 2.50, p = 0.01, are significant. A simple slopes analysis indicates that the probability of an error of omission increases significantly across trials during cycles 1, Exp(b) = 1.12, z = 3.22, $p < 0.01$, and 4, Exp(b) = 1.08, z = 2.45, p = 0.01, but not cycles 2 and 3, for participants in the Related Cues condition. None of the simple slopes for participants in the Unrelated Cues condition are significantly different than zero.

Similarly, the 3-way interaction between age group, cycle, and trial within cycle is not significant, F(3,6007) = 0.35, p = 0.79, but does yield significant parameter estimates for older adults in cycles 2, Exp(B) = 1.32, z = 2.85, $p < 0.01$, and 3, Exp(B) = 1.27, z = 2.45, p = 0.01. A simple slopes analysis indicates that young adults' omission rates increase significantly across trials only during cycle 1, Exp(b) = 1.07, z = 2.26, p = 0.02, but not for older adults, whose omission by trial slopes only approach significance in cycle 1, Exp(b) = 1.07, z = 1.71, p = 0.09, cycle 2, Exp(b) = 1.06, z = 1.64, p = 0.10, and cycle 4, Exp(b) = 1.07, z = 1.66, p = 0.10.

The significant 3-way interaction between age group, condition, and cycle indicates that the probability of an error of omission in a given recall block is different across experimental conditions and age groups, F(3,6007) = 2.73, p = 0.04. This effect appears to be driven by older adult participants in the Related Cues condition, whose odds of omitting answers during recall are low in cycles 1, Exp(b) = 0.11, SE = 0.04, 2, Exp(b) = 0.13, SE = 0.04, 3, Exp(b) = 0.13, SE = 0.04, and 4, Exp(b) = 0.09, SE = 0.03, compared to all other groups.

While the overall effect is not significant, F(3,6007) = 0.66, p = 0.58, there are significant parameter estimates in the cycle by trial interaction, particularly for cycles 2, Exp(B) = 0.80, z = -3.68, $p < 0.01$, and 3, Exp(B) = 0.82, z = -3.27, $p < 0.01$. A simple slopes analysis indicates that the probability of omission significantly increases across trials in cycles 1, Exp(b) = 1.05, z = 1.98, p = 0.05, and 4, Exp(b) = 1.07, z = 2.82, $p < 0.01$, but that the positive slopes in cycles 2, Exp(b) = 1.01, z = 0.53, p = 0.59, and 3, Exp(b) = 1.03, z = 1.21, p = 0.23, are not significantly different than zero.

Mixed evidence in support of differing omission by trial slopes by experimental condition is present in the insignificant trial by condition interaction, F(1,6007) = 2.05, p = 0.15, but significant parameter estimate, Exp(B) = 0.82, z = -3.00, $p < 0.01$. A simple slopes analysis indicates that the probability of errors of omission significantly increases across trials for participants in the Related Cues condition, Exp(b) = 1.12, z = 3.22, $p < 0.01$, but not the Unrelated Cues condition, Exp(b) = 1.02, z = 0.56, p = 0.57 (Figure 3.7).

The interaction between cycle and experimental condition is significant, F(1,6007) = 2.05, p = 0.03, indicating that differences in the probability of errors of omission across recall blocks is different between the Related Cues and Unrelated Cues conditions. An analysis of the marginal means indicates that overall probabilities of omitting responses during recall are stable across cycles 1, Exp(b) = 0.14, SE = 0.03, 2, Exp(b) = 0.18, SE = 0.04, 3, Exp(b) = 0.18, SE = 0.04, and 4, Exp(b) = 0.16, SE = 0.03, while the probabilities for participants in the Unrelated Cues condition are low in cycles 1, Exp(b) = 0.18, SE = 0.04, and 2, Exp(b) = 0.17, SE = 0.04, but increase in cycles 3, Exp(b) = 0.21, SE = 0.03, and 4, Exp(b) = 0.24, SE = 0.05.

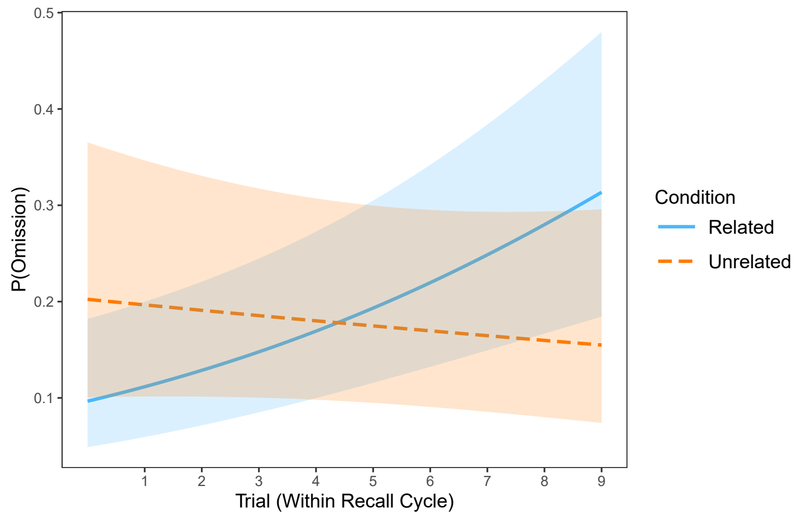

Figure 3.7: Simple slopes analysis estimating the probability of errors of omission across recall trials for Related and Unrelated Cues conditions.

The main effect of trial is significant, F(1,6007) = 10.68, $p < 0.01$, and the overall parameter estimate, Exp(B) = 1.17, z = 3.69, $p < 0.01$, indicates that the probability of errors of omission increases across trials, regardless of experimental condition or age group.

While the overall main effect of cycle is not significant, F(3,6007) = 1.20, p = 0.31, the individual parameter estimates for cycle 2, Exp(B) = 4.14, z = 4.34, $p < 0.01$, cycle 3, Exp(B) = 3.39, z = 3.70, $p < 0.01$, and cycle 4, Exp(B) = 2.53, z = 2.83, $p < 0.01$, are significant. A paired samples analysis on the estimated marginal means using Tukey's method indicates that differences in the probability of committing an error of omission in cycle 1, Exp(b) = 0.16, SE = 0.02, compared to cycles 3, Exp(b) = 0.19, SE = 0.03, p = 0.08, and 4, Exp(b) = 0.19, SE = 0.03, p = 0.07, approach significance.

Finally, the main effect of age group is significant, F(1,166.2) = 4.77, p = 0.03. An analysis of the estimated marginal means indicates that the probability of omitting an answer during cued recall is significantly lower for older adults, Exp(b) = 0.14, SE = 0.03, compared to young adults, Exp(b) = 0.22, SE = 0.04.

Errors of Commission vs. Other. Finally, I submitted these data to a logistic multilevel model examining errors of commission compared to all other responses during cued recall, using age group (Young vs. Older Adults), condition (Related Cues vs. Unrelated Cues), and recall block (or "cycle"; 1 - 4) as categorical predictors and trial within recall block (1 - 10) as a continuous predictor.

Interestingly, none of the overall effects, nor any of the parameter estimates, were significant, despite the overall model accounting for a large amount of information in the data, R^2_{SB} = 0.44. Both R (via `lme4`) and SAS (via `GLIMMIX`) were able to converge on similar solutions, and the outputs suggest that there were a sufficient number of observations to derive estimates for each parameter. Given that the fixed effects account for a low amount of information in the data, R^2_{SB} = 0.02, the two most likely explanations are 1) that relative probabilities of commission are constant across all fixed variables (particularly after accounting for random variations between participants and recall cycles) or 2) there are are variations in these relative probabilities of commission that are accounted for by variables outside of this experimental context.

3.2.4 Summary

The results of these analyses indicate that overall recall accuracy decreases across trials within recall blocks, but only when participants are experiencing output interference, i.e. when cue words belong to the same taxonomic category. This pattern is not stable across successive blocks of trials, however; participants in the Related Cues condition show heterogeneity in the recall by trial slopes across cycles (Figure 3.4). This pattern holds even for older adult participants in the Related Cues condition, where average recall accuracy is unexpectedly high compared to all other experimental groups.

For participants experiencing output interference in this task, decreases in correct recall across trials led to related increases in errors of omission, but not commission—again, with heterogeneity in slopes across the 4 blocks of recall trials for both young and older adults. Overall, the results indicate that, under conditions of output interference, correct recall of a given target word trades off with omission of a response during later trials in a given recall

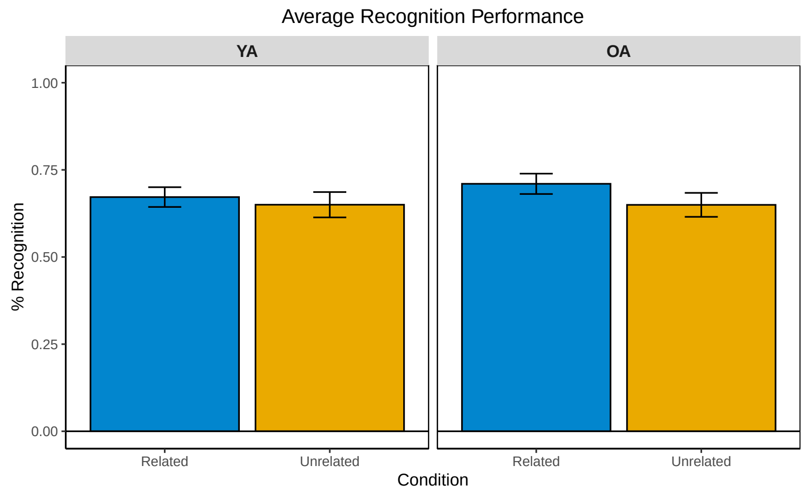

Figure 3.8: Average 4-alternative forced choice recognition memory performance by age group and cue-relatedness.

block. The probability of producing an incorrect item during cued recall (i.e. commission) did not reliably change across trials.

3.3 Recognition

Performance during the 4-alternative forced choice recognition test is important insofar as it provides the basis for participants' FOKs and RCJs during the task. A goal in developing this study was to produce recognition memory performance that is the same across age groups, experimental conditions, and trials, particularly when participants are tested on multiple word pairs whose cue words are semantically-related. Given this, I hypothesized that there would be no significant interactions or main effects in these 4AFC analyses.

3.3.1 Overall Means

After calculating mean 4AFC accuracy for each participant, I submitted these data to a 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA (Figure 3.8). Neither the main effects of Age Group, $F(1,147) = 0.77$, $p = 0.38$, and Condition $F(1,147) = 0.26$, $p = 0.61$, nor the interaction

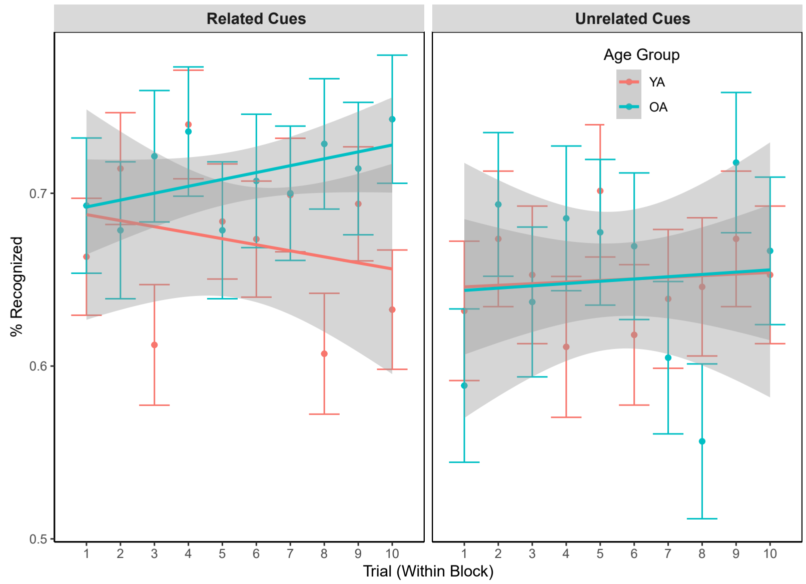

Figure 3.9: Average 4-alternative forced choice recognition memory performance by age group and cue-relatedness across blocked recognition trials.

between the two, $F(1{,}147) = 0.35$, $p = 0.55$, are significant. Young adults in the Related Cues condition have slightly higher recognition memory accuracy ($M = 0.67$, $SE = 0.03$) than young adults in the Unrelated Cues condition ($M = 0.65$, $SE = 0.04$), although this difference is not significant. This is similar for older adults, where those in the Related Cues condition have a slight, but insignificant advantage in recognition memory accuracy ($M = 0.71$, $SE = 0.03$) than those in the Unrelated Cues condition ($M = 0.65$, $SE = 0.03$).

I examined the evidence in favor of the null hypothesis of no Condition and Age Groups effects (H_0) against the alternative hypotheses of a main effect of Age Group (H_1), main effect of Condition (H_2), and interaction between the two (H_3). The Bayes Factor analysis indicates anecdotal to positive evidence in favor of the null hypothesis for all comparisons ($BF_{01} = 4.83$, $BF_{02} = 2.98$, $BF_{03} = 5.06$).

3.3.2 Summary

A simple test of group differences indicates that young and older adults performed similarly during a 4-alternative forced choice recognition memory task, regardless of whether participants experienced output interference during cued recall or not.

While it is not central to the hypotheses in this study, nor was it expected to show in the results, I performed an additional analysis to determine if recognition memory accuracy also shows signs of output interference during 4AFC (c.f. Criss et al., 2011). Recognition memory performance shows some changes across blocks of recognition trials with related cues (Figure 3.9), but a logistic GLM indicates that these changes are not significant (section A.2). Similarly, recognition memory performance is weakly related to previous memory outcomes during cued recall (section A.3), but with no significant influences from output interference states. Thus, overall recognition performance is relatively stable across trials.

3.4 Remember/Know/No Memory

3.4.1 Overall Means

The majority of studies employing Remember/Know/No Memory (R/K/N) judgments test mean differences in endorsement rates by averaging over conditions and analyzing these means using common general linear models, such as mixed-effects ANOVAs. I conducted such an analysis on these data (section A.4); however, the results are potentially problematic. In addition to the wide-spread violations of sphericity, a major concern in these ANOVAs is the normality of the observations. A Shapiro-Wilk test on the observations is significant ($p < 0.05$) for 8 out of the 12 cells in this analysis, suggesting that the assumption of normality is largely violated in this dataset. In order to alleviate this concern, I also submitted the data to a multinomial logistic regression, where the item-level responses (i.e. R/K/N) are the dependent variables. Here, "Remember", older adults, and the Related Cues conditions serve as the reference groups against all other levels in the Rating, Age Group, and Condition factors, respectively. Overall, the model provided a significantly greater fit to the data compared to a null (intercept) model, $\chi(6) = 111.80$, p $<$ 0.01.

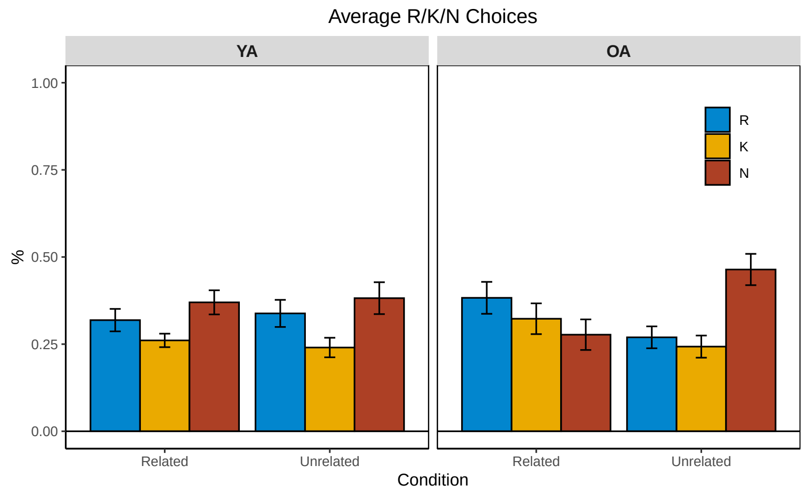

Figure 3.10: Average R/K/N judgment rates over age groups and experimental conditions.

Overall, young adults in the Related Cues condition are 1.22 times more likely to report "Remember" over "Know", Exp(B) = 0.82, z = -3.38, $p < 0.01$, but 1.16 times more likely to report "No Memory" than "Remember", Exp(B) = 1.16, z = 2.72, $p < 0.01$. Compared to young adults in the Related Cues condition, older adults in the Related Cues are 1.61 times more likely to report "Remember" compared to "No Memory", Exp(B) = 0.62, z = -5.47, $p < 0.01$; however, older adults in the Unrelated Cues condition are 2.44 times more likely to report "No Memory" compared to "Remember", Exp(B) = 2.44, z = 7.05, $p <$ 0.01. No other coefficients are significant.

3.4.2 R/K/N Endorsement Across Trials

I additionally submitted these data to a multinomial logistic regression with age group, condition, and trial within recall blocks (i.e. 10 trials over 4 blocks of recall trials) to assess whether R/K/N judgments change reliably across trials, and if so, whether change is different across experimental groups (Table 3.5). The "Remember" judgment is set as the judgment against which "Know" and "No Memory" judgments are compared, with young adults in the Related Cues condition being set as the reference group for the predictors. Overall, the full model accounts for a significantly greater amount of information in the

	Dependent variable:	
	Know	No Memory
	(1)	(2)
Intercept	$-\mathbf{0.201}^{*}$	$\mathbf{0.148}^{*}$
	(0.060)	(0.055)
Age (OA)	0.031	$-\mathbf{0.472}^{*}$
	(0.087)	(0.086)
Condition (UC)	−0.140	−0.027
	(0.092)	(0.083)
Age (OA) x Condition (UC)	0.207	$\mathbf{0.893}^{*}$
	(0.137)	(0.127)
-2LL	12,593.92	
AIC	12,609.92	
$\Delta\chi^2$ (against null)	111.80	$p < 0.01$
Note:		$^{+}$p<0.1; *p<0.05

Table 3.4: Multinomial logistic model predicting R/K/N (in log-odds) from age group and experimental condition.

data compared to the null (intercept) model, $\chi^2(14)$ = 138.28.

For “No Memory” judgments, the estimate for judgment endorsement across trials is significant for participants in the Unrelated Cues condition, Exp(B) = 0.92, z = -2.93, $p < 0.01$, indicating that the probability of endorsing *N* significantly decreases across trials relative to endorsing *R*; however, this slope is not significant for “Know” responses, Exp(B) = 0.99, z = -0.42, p = 0.68.

The parameter estimate for “No Memory” judgments is also significant for older adults in the Unrelated Cues condition, Exp(B) = 2.19, z = 3.31, $p < 0.01$, indicating that the *N* judgment is approximately 2.19 times more likely for older adult participants in the Unrelated Cues condition than for participants in the reference group, i.e. young adults in the Related Cues condition. Similarly, older adult participants in the Unrelated Cues condition are more 1.67 times more likely to endorse “Know” judgments compared to the reference group, Exp(B) = 1.67, z = 2.02, p = 0.04.

The overall probability of endorsing “No Memory” significantly increases over trials

	Dependent variable:	
	Know	No Memory
	(1)	(2)
Intercept	−0.161	−0.115
	(0.107)	(0.103)
Age (OA)	−0.214	**−0.488***
	(0.160)	(0.162)
Condition (Unrelated)	−0.078	**0.357***
	(0.169)	(0.155)
Trial (Within Cycle)	−0.010	**0.058***
	(0.021)	(0.019)
Age (OA) x Condition (Unrelated)	**0.510***	**0.786***
	(0.253)	(0.237)
Age (OA) x Trial	**0.056***	0.004
	(0.031)	(0.030)
Condition (Unrelated) x Trial	−0.013	**−0.084***
	(0.032)	(0.029)
Age (OA) x Condition (Unrelated) x Trial	−0.069	0.023
	(0.048)	(0.044)
-2LL	12,705.72	
AIC	12,599.44	
$\Delta\chi^2$ (against null)	138.28	$p < 0.01$
Note:		$^{+}$p<0.1; *p<0.05

Table 3.5: Multinomial logistic model predicting R/K/N (in log-odds) from age group, experimental condition, and trial within recall blocks.

(compared to "Remember" judgments), Exp(z) = 1.06, z = 3.02, $p < 0.01$, while the probability of endorsing "Know" numerically, but not significantly, declines, Exp(B) = 0.99, z = -0.46, p = 0.65.

Finally, the significant "No Memory" parameter estimate for older adults, Exp(B) = 0.61, z = -3.01, $p < 0.01$, indicates that older adults are less likely to report states of unknowing during recall compared to young adults. No other estimates are significant.

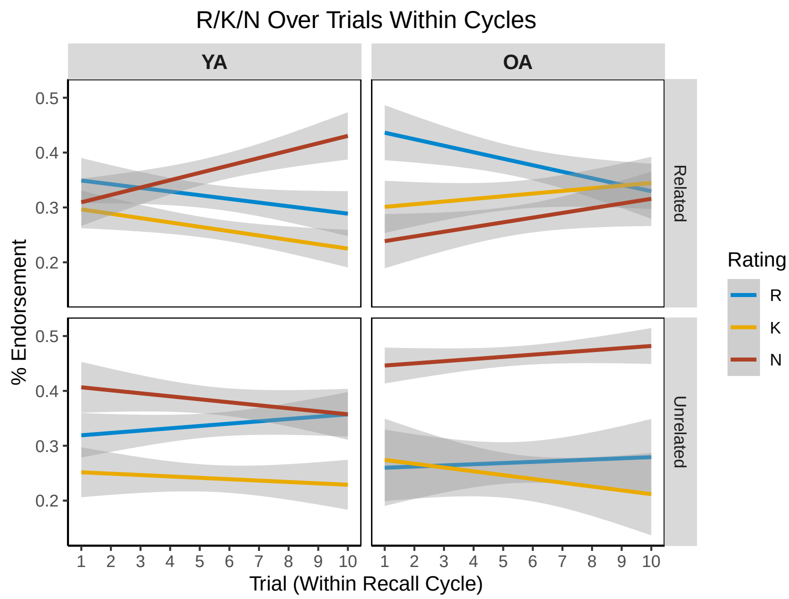

Figure 3.11: Average R/K/N judgment rates over trials within recall cycles.

3.4.3 Multilevel Models

As with recall and recognition outcomes, the multilevel modeling framework is a prime candidate for simultaneously accounting for group-, participant-, and cycle-level variation in R/K/N judgment behavior. Here, I describe 3 separate binomial models using a logit link function: R vs. $\{K,N\}$, K vs. $\{R,N\}$, and N vs. $\{R,K\}$. Like the multilevel model for recall outcomes, I dichotomized the 3 separate judgments and submitted these data to separate models with logit link functions on the dependent variable (Snijders & Bosker, 2011). The parameters of these models can be interpreted as the likelihood of endorsing a given judgment against all other judgments. Age group (Young vs. Older Adults), experimental condition (Related Cues vs. Unrelated Cues), recall cycle (1 - 4), trial number within cycle (1 - 10), and item-level recall accuracy were included as main effects in each of the 3 multilevel models.

Random Components. Before specifying full models for each of the comparisons, I performed model comparisons on various random components using single-term dropout to determine whether random intercepts across participants and recall trial blocks account for a significant amount of information accounted for by the 3 models. For the "Remember" model, excluding both random effects of cycle, $\chi^2(1)$ = 6.21, p = 0.01, and participant, $\chi^2(1)$ = 1059.51, $p < 0.01$, results in significant losses of information explained by the model. Similarly, removing the random effects of cycle, $\chi^2(1)$ = 8.85, $p < 0.01$, and participant, $\chi^2(1)$ = 1441.50, $p < 0.01$, results in significant losses of information in the data accounted for by the model. For the "Know" model, however, only the removal of the random effect of participants is significant, $\chi^2(1)$ = 654.74, $p < 0.01$, but not when the random effect of cycle is removed, $\chi^2(1)$ = 0.11, p = 0.744. Despite this, I retained random effects of both participants and cycles for all models.

Model Reduction. To help reduce the complexity of the multilevel models, I used single-term dropout to determine which terms can be excluded without significant loss of information explained by the model. I started by excluding the highest-level interactions and working down through successively lower-level interactions. If excluding a term did not result in significant loss of information in a model, then that term was deleted from all 3 models, and single-term deletion analyses continued using the new, compact models.

Likelihood ratio tests on models that exclude the 5-way interaction (i.e. Age Group x Condition x Cycle x Recall) did not reach significance for the "Remember" model, $\chi^2(3)$ = 7.46, p = 0.06, the "Know" model, $\chi^2(3)$ = 2.58, p = 0.46, and the "No Memory" model, $\chi^2(3)$ = 0.25, p = 0.97. Despite the test approaching significance for the "Remember" model, I decided to exclude the term from all 3 models.

The 4-way interaction between condition, cycle, trial, and recall accuracy was also not significant for any of the 3 models, $\chi^2_R(3)$ = 2.65, p = 0.45, $\chi^2_K(3)$ = 0.94, p = 0.82, $\chi^2_N(3)$ = 2.09, p = 0.55, nor was the 4-way interaction between age group, condition, cycle, and recall accuracy, $\chi^2_R(3)$ = 3.18, p = 0.36, $\chi^2_K(3)$ = 2.17, p = 0.54, $\chi^2_N(3)$ = 3.75, p = 0.29. The last term dropped from the model without significant loss of information was the 4-way interaction between age group, condition, cycle, and cycle trial, $\chi^2_R(3)$ = 0.65, p =

0.89, $\chi^2_K(3) = 2.21$, $p = 0.53$, $\chi^2_N(3) = 1.47$, $p = 0.69$. No other 4-way interactions could be removed without significant loss of information accounted for in at least one model.

Model fit did not significantly differ between the new reduced models and models with all possible terms included, $\chi^2_R(12) = 13.96$, $p = 0.30$, $\chi^2_K(12) = 7.89$, $p = 0.79$, $\chi^2_N(12) = 7.56$, $p = 0.82$.

Table 3.6: *F*-table for the multilevel models predicting "Remember" (Model 1), "Know" (Model 2), and "No Memory" (Model 3) judgments against all others.

	Remember (1)		Know (2)		No Memory (3)	
	df	*F*	*df*	*F*	*df*	*F*
Age Group	1, 242.6	0.60	1, 223.7	0.20	1, 501.2	1.59
Condition	1, 244.5	**4.68***	1, 224.4	1.05	1, 543	3.00^{+}
Cycle	3, 5987	0.59	3, 5987	0.03	3, 5987	0.24
Trial (Within Cycle)	1, 5987	1.07	1, 5987	3.01^{+}	1, 5987	0.37
Recall	1, 5987	**400.44***	1, 5987	2.52	1, 5987	**211.01***
Age x Condition	1, 243.5	0.62	1, 224.6	0.47	1, 537.3	0.08
Age x Cycle	3, 5987	1.27	3, 5987	0.34	3, 5987	0.52
Condition x Cycle	3, 5987	0.89	3, 5987	1.13	3, 5987	2.48^{+}
Age x Trial	1, 5987	0.41	1, 5987	0.30	1, 5987	0.61
Condition x Trial	1, 5987	**7.36***	1, 5987	0.92	1, 5987	3.01^{+}
Cycle x Trial	3, 5987	0.69	3, 5987	1.24	3, 5987	1.33
Age x Recall	1, 5987	2.61^{+}	1, 5987	3.40^{+}	1, 5987	3.22^{+}
Condition x Recall	1, 5987	**13.46***	1, 5987	0.31	1, 5987	3.11^{+}
Cycle x Recall	3, 5987	0.58	3, 5987	0.60	3, 5987	0.09
Trial x Recall	1, 5987	**4.27***	1, 5987	0.03	1, 5987	2.48
Age x Condition x Cycle	3, 5987	0.73	3, 5987	1.69	3, 5987	0.38
Age x Condition x Trial	1, 5987	0.10	1, 5987	**6.25***	1, 5987	1.33
Age x Cycle x Trial	3, 5987	0.61	3, 5987	0.11	3, 5987	0.55
Condition x Cycle x Trial	3, 5987	0.86	3, 5987	0.41	3, 5987	1.05
Age x Condition x Recall	1, 5987	0.02	1, 5987	1.27	1, 5987	0.59
Age x Cycle x Recall	3, 5987	1.11	3, 5987	**4.39***	3, 5987	0.10
Condition x Cycle x Recall	3, 5987	1.23	3, 5987	1.52	3, 5987	2.02
Cycle x Trial x Recall	3, 5987	0.47	3, 5987	0.55	3, 5987	0.53
Age x Condition x Trial x Recall	2, 5987	**3.60***	2, 5987	0.87	2, 5987	0.08
Age x Cycle x Trial x Recall	3, 5987	1.39	3, 5987	**2.82***	3, 5987	0.79

Note: $^{+}$p<0.1; *p<0.05

Table 3.7: Parameter estimates for the multilevel models predicting "Remember" (Model 1), "Know" (Model 2), and "No Memory" (Model 3) judgments against all others. Results are reported in log-odds.

	Dependent variable:		
	Remember	Know	No Memory
	(1)	(2)	(3)
Intercept	**−2.056***	−0.472^{+}	−0.341
	(0.393)	(0.281)	(0.332)
Age Group (OA)	0.384	**−1.184***	0.335
	(0.594)	(0.452)	(0.517)
Condition (Unrelated)	**−1.384***	−0.356	**1.100***
	(0.576)	(0.391)	(0.483)
Cycle 2	0.040	−0.417	0.095
	(0.454)	(0.312)	(0.327)
Cycle 3	0.219	−0.242	0.032
	(0.459)	(0.318)	(0.336)
Cycle 4	−0.196	**−0.674***	0.518
	(0.466)	(0.324)	(0.333)
Trial (Within Cycle)	−0.110	−0.066	**0.091***
	(0.068)	(0.043)	(0.045)
Recall	**3.276***	**−0.902***	**−4.297***
	(0.436)	(0.352)	(0.983)
Age (OA) x Condition (Unrelated)	−0.702	0.504	0.397
	(0.731)	(0.532)	(0.667)
Age (OA) x Cycle 2	0.151	**0.986***	−0.699
	(0.679)	(0.494)	(0.504)
Age (OA) x Cycle 3	−0.736	**1.417***	−0.688
	(0.684)	(0.487)	(0.501)
Age (OA) x Cycle 4	0.012	**1.768***	**−1.403***
	(0.697)	(0.492)	(0.505)
Condition (Unrelated) x Cycle 2	0.304	0.490	−0.541
	(0.571)	(0.393)	(0.454)
Condition (Unrelated) x Cycle 3	0.912	−0.068	−0.460
	(0.562)	(0.400)	(0.456)
Condition (Unrelated) x Cycle 4	0.421	−0.157	−0.116
	(0.580)	(0.407)	(0.460)
Age (OA) x Trial	−0.074	**0.187***	−0.085
	(0.104)	(0.066)	(0.066)
Condition (Unrelated) x Trial	**0.251***	0.016	**−0.175***
	(0.080)	(0.054)	(0.060)
Cycle 2 x Trial	0.036	0.018	−0.032
	(0.089)	(0.058)	(0.060)
Cycle 3 x Trial	−0.057	0.008	0.019

	(0.091)	(0.058)	(0.061)
Cycle 4 x Trial	0.018	0.013	−0.030
	(0.089)	(0.058)	(0.059)
Age (OA) x Recall	−1.156^{+}	**1.587***	1.251
	(0.654)	(0.542)	(1.117)
Condition (Unrelated) x Recall	**1.841***	−0.615	−0.224
	(0.546)	(0.424)	(1.039)
Cycle 2 x Recall	−0.270	0.421	−0.351
	(0.591)	(0.483)	(1.353)
Cycle 3 x Recall	−0.690	0.433	1.019
	(0.577)	(0.477)	(1.183)
Cycle 4 x Recall	−0.561	**1.285***	−0.333
	(0.589)	(0.481)	(1.396)
Trial x Recall	**0.253***	−**0.206***	0.028
	(0.115)	(0.081)	(0.124)
Age (OA) x Condition (Unrelated) x Cycle 2	0.682	−**0.775***	0.194
	(0.463)	(0.371)	(0.467)
Age (OA) x Condition (Unrelated) x Cycle 3	0.326	−0.622^{+}	0.380
	(0.475)	(0.375)	(0.462)
Age (OA) x Condition (Unrelated) x Cycle 4	0.360	−0.621^{+}	0.447
	(0.480)	(0.377)	(0.462)
Age (OA) x Condition (Unrelated) x Trial	−0.160^{+}	−0.074	0.093
	(0.094)	(0.060)	(0.059)
Age (OA) x Cycle 2 x Trial	0.064	−0.120	0.089
	(0.134)	(0.085)	(0.084)
Age (OA) x Cycle 3 x Trial	0.239^{+}	−**0.207***	0.041
	(0.133)	(0.083)	(0.083)
Age (OA) x Cycle 4 x Trial	0.075	−0.158^{+}	0.129
	(0.134)	(0.083)	(0.082)
Condition (Unrelated) x Cycle 2 x Trial	−0.100	−0.004	0.111
	(0.081)	(0.063)	(0.078)
Condition (Unrelated) x Cycle 3 x Trial	−0.115	0.043	0.083
	(0.082)	(0.064)	(0.077)
Condition (Unrelated) x Cycle 4 x Trial	−0.039	0.053	0.003
	(0.083)	(0.065)	(0.077)
Age (OA) x Condition (Unrelated) x Recall	−0.092	0.581	−0.881
	(0.651)	(0.513)	(1.125)
Age (OA) x Cycle 2 x Recall	0.429	−1.306^{+}	0.322
	(0.849)	(0.698)	(1.530)
Age (OA) x Cycle 3 x Recall	1.478^{+}	−**1.919***	−0.213
	(0.851)	(0.695)	(1.387)
Age (OA) x Cycle 4 x Recall	0.819	−**2.401***	0.531
	(0.862)	(0.701)	(1.535)
Condition (Unrelated) x Cycle 2 x Recall	−0.797	0.581	−0.091
	(0.490)	(0.379)	(0.785)

Condition (Unrelated) x Cycle 3 x Recall	−0.807	**0.808***	−**1.638***
	(0.496)	(0.389)	(0.828)
Condition (Unrelated) x Cycle 4 x Recall	−0.794	0.489	0.355
	(0.501)	(0.392)	(0.819)
Cycle 2 x Trial x Recall	0.004	−0.041	0.358
	(0.108)	(0.086)	(0.289)
Cycle 3 x Trial x Recall	0.165	−0.140	0.170
	(0.109)	(0.087)	(0.281)
Cycle 4 x Trial x Recall	0.104	−0.130	−0.012
	(0.110)	(0.089)	(0.373)
Age (YA) x Condition (Related) x Trial x Recall	−0.202	**0.307***	−0.241
	(0.142)	(0.103)	(0.283)
Age (OA) x Condition (Related) x Trial x Recall	−0.086	0.093	0.003
	(0.089)	(0.070)	(0.117)
Age (YA) x Condition (Unrelated) x Trial x Recall	−**0.398***	**0.302***	−0.318
	(0.148)	(0.109)	(0.309)
Age (OA) x Cycle 2 x Trial x Recall	−0.163	0.244^{+}	−0.483
	(0.166)	(0.130)	(0.327)
Age (OA) x Cycle 3 x Trial x Recall	−**0.339***	**0.352***	−0.284
	(0.167)	(0.132)	(0.315)
Age (OA) x Cycle 4 x Trial x Recall	−0.204	**0.304***	−0.176
	(0.169)	(0.133)	(0.406)
σ^2 (Recall Cycle)	<0.01	<0.01	<0.01
τ^2 (Participants)	1.27	1.05	1.51
-2LL	−2,261.507	−3,131.709	−2,281.195
R^2_{SB} (Fixed)	0.40	0.03	0.49
R^2_{SB} (Overall)	0.60	0.27	0.70
Note:			$^{+}$p<0.1; *p<0.05

“Remember” vs. Other. The first model in the series examined the item-level likelihood of endorsing “Remember”—a judgment that is thought to represent a conscious recollective experience during memory retrieval (Yonelinas, 2002)—against all other judgments. The overall model accounted for a large amount of information in the model, R^2_{SB} = 0.60, as did the fixed effects, R^2_{SB} = 0.40.

While the 4-way interaction between age group, recall cycle, trial (within cycle), and recall accuracy is not significant in the aggregate, F(3,5987) = 1.39, p = 0.24, older adults' trial by *R* endorsement slope for correctly-recalled items during cycle 3 is significant, Exp(B) = 0.71, z = -2.03, p = 0.04. Even with the significant parameter estimate, a simple slopes analysis indicates that none of the slopes are significantly different than zero.

The 4-way interaction between age group, condition, cycle, and recall accuracy is significant, F(2, 5987) = 3.60, p = 0.02, with a significant parameter estimate for the *R* endorsement by trial slope for correctly-recalled items for young adults in the Unrelated Cues condition, Exp(B) = 0.67, z = -2.69, p = 0.01. A simple slopes analysis indicates that the probability of endorsing *R* for unrecalled items significantly decreases across trials for young adults in the Related Cues condition, Exp(b) = 0.83, z = -2.38, p = 0.02, but significantly increases for unrecalled items for young adults in the Unrelated Cues Condition Exp(b) = 1.40, z = 2.68, p = 0.01. The simple slopes for unrecalled items given by older adults in the Related Cues condition, Exp(b) = 0.85, z = -1.69, p = 0.09, and for recalled items given by older adults in the Unrelated Cues condition, Exp(b) = 1.17, z = 1.90, p = 0.06, approach, but do not reach, significance.

Interestingly, none of the 3-way interactions are significant, and the next significant term is the trial by recall accuracy interaction, F(1, 5987) = 4.2, p = 0.04, and its associated parameter estimate, Exp(B) = 1.29, z = 2.21, p = 0.03. A simple slopes analysis indicates that the effect is driven by the difference between the positive slope for recalled items, Exp(b) = 1.02, z = 0.42, p = 0.67, and a negative slope for unrecalled items, Exp(b) = 0.94, z = -1.14, p = 0.26, although neither of the slopes are significantly different than zero.

The interaction between experimental condition and recall accuracy is also significant, F(1, 5987) = 13.46, $p < 0.01$, as well as the parameter estimate for participants in the Unrelated Cues condition who correctly recall an item, Exp(B) = 6.30, p = 3.37, $p < 0.01$. A paired-samples test using Tukey's method indicates that the probability of participants in the Related Cues condition endorsing an *R* response after a failed recall attempt, Exp(b) = 0.09, SE = 0.01, is significantly different from participants in the same condition that correctly recall an item, Exp(b) = 1.99, SE = 0.04, $p < 0.01$, participants in the Unrelated Cues condition who also fail to recall an item, Exp(b) = 0.08, SE = 0.01, p = 0.05, and participants in the Unrelated Cues condition who correctly recall an item, Exp(b) = 2.63, SE = 0.04, $p < 0.01$. The probability of endorsing *R* is significantly lower for participants in the Unrelated Cues condition during an unsuccessful recall attempt compared to participants in the same condition who correctly recall an item, $p < 0.01$, and participants in the Related Cues condition who correctly recall an item, $p < 0.01$. There is no significant difference between the two experimental conditions when an item is correctly recalled, p = 0.65.

The interaction between experimental condition and trial within a recall block is significant, F(1, 5987) = 7.36, $p < 0.01$, with a positive and significant parameter estimate for participants in the Unrelated Cues condition, Exp(B) = 1.28, z = 3.14, $p < 0.01$. A simple slopes analysis (Figure 3.12) indicates that the relative probability of endorsing "Remember" significantly decreases across trials for participants in the Related Cues condition, Exp(b) = 0.91, z = -2.52, p = 0.01, but not for participants in the Unrelated Cues condition, Exp(b) = 1.04, z = 0.75, p = 0.46, where the slope is not significantly different than zero. No other interactions are significant.

The main effect of recall accuracy is significant, F(1, 5987) = 400.44, $p < 0.01$, and the significant parameter estimate, Exp(B) = 26.58, z = 7.51, $p < 0.01$, reflects the significantly lower odds of endorsing the "Remember" judgment when recall is unsuccessful, Exp(b) = 0.07, SE = 0.01, compared to when recall is successful, Exp(b) = 2.29, SE = 0.03.

Finally, the main effect of experimental condition is significant, F(1, 244.5) = 4.68, p =

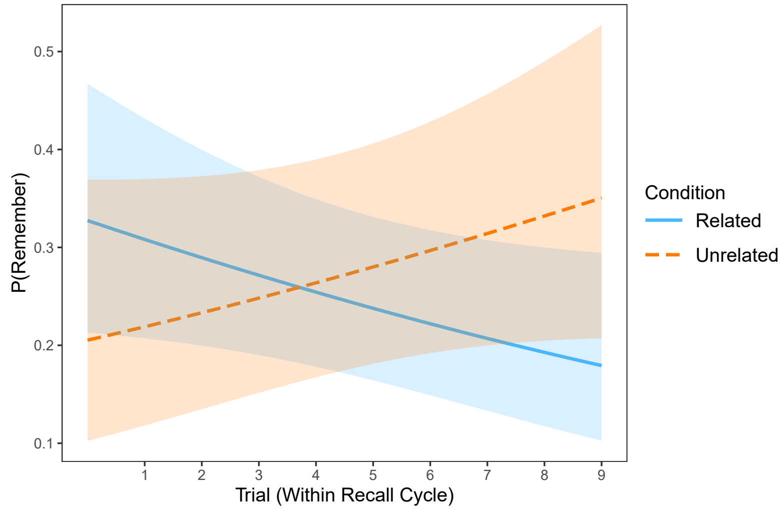

Figure 3.12: Simple slopes analysis estimating the probability of endorsing "Remember" across recall trials for Related and Unrelated Cues conditions.

0.03, with a significant parameter estimate for participants in the Unrelated Cues condition, Exp(B) = 0.25, z = -2.40, p = 0.02. Participant in the Related Cues condition are more likely to endorse *R* judgments, Exp(b) = 0.43, SE = 0.03, than are participants in the Unrelated Cues condition, Exp(b) = 0.35, SE = 0.03, likely reflecting significantly greater recall performance for older adult participants in the Related Cues condition.

"Know" vs. Other. Data that compares endorsement of "Know" to all other judgments was submitted to a second binomial multilevel model, but with lower model fit: The full model accounts for a moderate amount of information in the data, R^2_{SB} = 0.27, while the fixed effects only account for a small amount of information, R^2_{SB} = 0.03. Despite this, several of the terms in the model are significant.

The first significant term is the 4-way interaction between age, recall cycle, trial (within recall cycle), and recall outcome, F(3, 5987) = 2.82, p = 0.04, with significant parameter estimates for *K* endorsement over trials for older adults who correctly recall an item during cycles 3, Exp(B) = 1.42, z = 2.67, p = 0.01, and 4, Exp(B) = 1.36, z = 2.29, p = 0.02. A simple slopes analysis indicates that the only slope significantly different than zero is for older adults during the first cycle in the event of a incorrect recall, Exp(b) = 1.11, z = 2.18,

p = 0.03.

The 4-way interaction between age group, condition, recall accuracy, and trial is not significant, F(2, 5987) = 0.87, p = 0.42, but does yield significant parameter estimates for young adults in the Related Cues condition, Exp(B) = 1.35, z = 2.98, $p < 0.01$, and young adults in the Unrelated Cues condition, Exp(B) = 1.35, z = 2.76, p = 0.01, but only for items that are correctly recalled. A simple slopes analysis indicates that only the slope for older adults in the Related Cue condition for incorrect recall is significantly different than zero, Exp(b) = 1.16, z = 2.23, p = 0.03, indicating that older adults' endorsement of the "Know" judgment increases across trials, but only when recall is unsuccessful.

The 3-way interaction between age group, cycle, and recall accuracy is significant, F(3, 5987) = 4.39, $p < 0.01$, with significant parameter estimates for cycles 3, Exp(B) = 0.30, z = -2.76, p = 0.01, and 4, Exp(B) = 0.09, z = -3.42, $p < 0.01$, with the estimate for cycle 2 only approaching significance, Exp(B) = 0.27, z = -1.87, p = 0.06. A paired comparison analysis using Tukey's method indicates that the effect is driven by young adults' relatively higher odds of endorsing *K* during a recall failure in the first cycle, Exp(b) = 0.40, SE = 0.03, compared to the last cycle, Exp(b) = 0.23, SE = 0.02, p = 0.02, and to young adults in the first cycle that correctly recall an item, Exp(b) = 0.19, SE = 0.03, $p < 0.01$. No other paired comparisons are significant.

The 3-way interaction between age group, condition, and trial within recall block is also significant, but with an insignificant parameter estimate for older adults in the Unrelated Cues condition, Exp(B) = 0.93, z = -1.23, p = 0.22. A simple slopes analysis (Figure 3.13) indicates that the effect is driven by increases in "Know" endorsement across trials for older adult participants in the Related Cues condition, Exp(b) = 1.07, z = 1.76, p = 0.08, although this effect only approaches significance. All other simple slopes were not significant, nor did they approach significance.

While the interaction between age group and recall only approaches significance, F(1, 5987) = 3.40, p = 0.07, the parameter estimate is significant, Exp(B) = 4.90, z = 2.93, $p <$

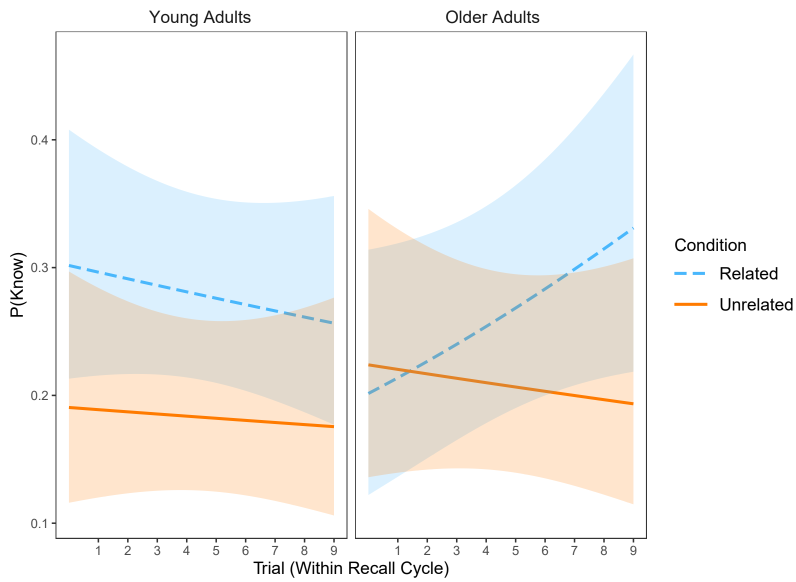

Figure 3.13: Simple slopes analysis estimating the probability of endorsing "Know" across recall trials for young and older adults in the Related and Unrelated Cues conditions.

0.01. An analysis of the marginal means using Tukey's paired comparison method indicates that young adults are significantly more likely to endorse the "Know" judgment when items are not recalled, Exp(b) = 0.33, SE = 0.02, compared to when items are recalled, Exp(b) = 0.23, SE = 0.02, $p < 0.01$; however, older adults are equally likely to endorse "Know" judgments when recall fails, Exp(b) = 0.32, SE = 0.03, as they are when recall is successful, Exp(b) = 0.32, SE = 0.03, p = 0.99. No other paired comparisons are significant.

None of the main effects are significant, although the parameter estimates for recall, Exp(B) = 0.41, z = -2.56, p = 0.01, cycle 4, Exp(B) = 0.51, z = -2.08, p = 0.04, and older adults, Exp(B) = 0.31, z = -2.62, p = 0.01, are significant. Paired comparison analyses on the individual main effects (using Tukey's method) only yield one significant comparison that indicates that participants are more likely to endorse *K* during incorrect recall, Exp(b) = 0.33, SE = 0.02, than when recall is successful, Exp(b) = 0.27, SE = 0.02, p = 0.01. None of the paired comparisons for *K* endorsement across age groups and recall cycles are significant.

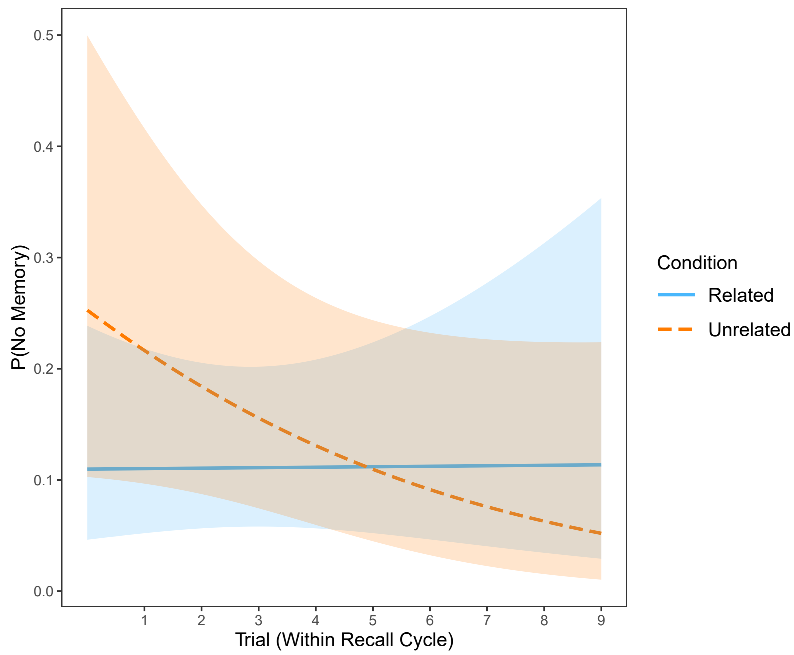

Figure 3.14: Simple slopes analysis estimating the probability of endorsing "No Memory" across recall trials for the Related and Unrelated Cues conditions.

"No Memory" vs. Other. Lastly, I estimated a model examining trial-level endorsement of the "No Memory" judgment against all others using the terms specified for the previous two models. Overall, the fixed effects, R^2_{SB} = 0.49, and the full model, R^2_{SB} = 0.70, account for a large amount of the information in the data.

Interestingly, only the main effect of recall accuracy is significant in the aggregate, F(1, 5987) = 211.01, with a corresponding significant parameter estimate for correct recall trials, Exp(B) = 0.01, z = -4.37, p < 0.01, reflecting a significantly greater probability of endorsing N when cued recall fails, Exp(b) = 1.42, SE = 0.03, compared to when it is successful, Exp(b) = 0.02, SE = 0.01, $p < 0.01$.

Some effects only approach significance in the aggregate, but have significant parameter estimates. For example, the condition by trial interaction, F(1, 5987) = 3.01, p = 0.08, has a significant estimate for participants in the Unrelated Cues condition, Exp(B) = 0.83, z = -2.91, p < 0.01. A simple slopes analysis (Figure 3.14) indicates that the probability of endorsing "No Memory" decreases across trials for participants in the Unrelated Cues

Condition, Exp(b) = 0.88, z = -1.89, p = 0.06, although this slope is not significantly different than zero. The corresponding slope for participants in the Related Cues condition is flat and non-significant, Exp(b) = 1.01, z = 0.10, p = 0.92. No other effects are significant.

3.4.4 Summary

During cued recall, both young adults and older adults in the Unrelated Cues condition reported fewer states of familiarity compared to recollection or no memory at all. For older adults in the Related Cues condition, however, experiences of familiarity exceeded those of having no memory, although there were a greater number of reported states of recollection than either of the other two experiences. This is undoubtedly connected to significantly greater cued recall performance for older adults in the Related Cues condition.

Importantly, when accounting for differences at the level of the participant and each recall block, the probability of reporting a recollective experience during recall significantly decreases with each successive recall trial within a block, but only when experiencing output interference (i.e. being cued with semantically-related words during a cued recall test). Familiarity only significantly changed across trials for older adults in the Related Cues condition, who were more likely to endorse "Know" judgments with increases output interference, although this slope only approaches significance. Finally, participants in the Unrelated Cues condition showed a decreased likelihood of endorsing "No Memory" across trials within a recall block, but this too only approached significance. Thus, only recollection reliably changed with output interference in this task.

3.5 Post-recognition Confidence Judgments

Post-recognition confidence judgments (RCJs) are given immediately after a recognition trial and provide a comparison to metacognitive monitoring and judgments that occur with respect to a prospective recognition test (FOKs). Because recognition memory is not affected by output interference or age in this task (c.f. Section 3.3), RCJs are also not

		RCJ x Recog. Gamma	
Age	Condition	*M*	*SE*
YA	Related	0.66	*0.05*
	Unrelated	0.60	*0.06*
OA	Related	0.65	*0.04*
	Unrelated	0.66	*0.05*

Table 3.8: Average gamma correlation estimates for RCJs across age groups and conditions.

expected to vary as a function of age or experimental condition.

This section will only detail analyses that explore trends in overall RCJs and relationships between RCJs and recognition memory performance. These judgments will be used to compare prospective (FOKs) and retrospective (RCJs) metacognitive experiences during 4AFC in the next section (3.6).

3.5.1 Overall Means

I submitted participant-level RCJ means to a 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA. Neither the main effects of Age Group, F(1,147) = 0.51, p = 0.48, and Condition F(1,147) = 0.19, p = 0.66, nor the interaction between the two, F(1,147) = 0.42, p = 0.52, are significant. A Bayes Factor analysis indicates positive evidence in favor of the null hypothesis (H_0) compared to models with effects of Age Group (H_1), Condition (H_2), or the interaction between the two (H_3): BF_{01} = 5.41, BF_{02} = 3.09, BF_{03} = 4.98.

3.5.2 Relative Accuracy

I computed Goodman-Kruskal gamma correlations (L. A. Goodman & Kruskal, 1963) between individual memory outcomes and item-level RCJs (Table 3.8). These data were then submitted to a 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA to determine if post-recognition confidence accuracy is similar between age groups and experimental conditions. Overall, neither the interaction between Age Group and Condition, F(1,146) = 0.40, p = 0.53, nor the separate main effects of Age Group, F(1,146) = 0.27, p = 0.60, and Condition, F(1,146) = 0.14, p = 0.71, are significant. A Bayes Factor analysis indicates positive evidence in favor of a null (i.e. no group differences) model against the interaction, BF_{01} = 4.98, and the separate main effects of Age Group, BF_{02} = 5.32, and Condition, BF_{03} = 5.24.

3.5.3 Summary

As expected, the mean gamma correlation between recognition performance and RCJs is relatively high for all participants, M = 0.65, SE = 0.02, and does not significantly differ between age groups or experimental conditions. Despite previous findings showing age-related deficits in RCJ resolution (Dodson et al., 2007), a null effect for age in this study is concordant with recent evidence suggesting that age-related differences in RCJ resolution are resolved when memory performance is equated (Hertzog et al., 2021).

3.6 Feelings-of-knowing

The most critical analyses in this study involve feelings-of-knowing (FOKs), specifically if FOKs reported by participants after cued recall trials track later recognition memory test outcomes, and if the relationship between FOKs and recognition memory are influenced by memory interference during recall. These analyses will indicate whether participants' FOKs reflect memory experiences in which output interference blocks access to non-criterial recollection during the judgment process (retrieval-suppression account) or not (competition account), and whether interference states differentially affect FOKs in older and young adults.

3.6.1 Overall Means

I first averaged FOK judgments within participants and then submitted these means to a 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs.

Unrelated Cues) between-subjects ANOVA (Table 3.1). Only the main effect of Age Group is significant, F(1,147) = 4.41, p = 0.04, $\eta_p^2 = 0.01$, where older adults report higher FOKs on average, M = 60.16, SE = 2.66, compared to young adults, M = 57.05, SE = 2.37. Neither the main effect of Condition, F(1,147) = 0.01, p = 0.97, nor the interaction between Age Group and Condition, F(1,147) = 3.78, p = 0.053, are significant, though the latter approached significance. A post-hoc analysis using Tukey's HSD indicates that the difference in mean FOKs between the Related and Unrelated Cues conditions is marginally significant for older adults, M_{RC} = 67.16, SE_{RC} = 3.65, M_{UC} = 53.16, SE_{UC} = 3.88, p = 0.05, but not young adults, M_{RC} = 57.12, SE_{RC} = 3.09, M_{UC} = 56.98, SE_{UC} = 3.60, p = 0.99. These differences in average FOK judgments are likely driven by high recall performance for older adults in the Related Cues condition compared to all other groups, even in the absence of a significant interaction. No other pairwise contrasts are significant.

A Bayes Factor analysis comparing evidence in support of the null hypothesis (i.e. no group differences; H_0) compared to models with the main effect of Age Group (H_1), main effect of Condition (H_2), and the interaction between the two (H_3), indicates marginal to positive evidence in favor of the null (BF_{01} = 3.65, BF_{02} = 1.50, BF_{03} = 1.29). The results of the Bayes Factor analysis, along with the ANOVA, suggest that there are no mean differences in FOK judgments across age groups or experimental conditions, although the evidence in support of this null hypothesis is anecdotal.

Across Recall Outcomes. I additionally examined mean FOKs by the outcome of the cued recall trial (c.f. Krinsky & Nelson, 1985), specifically for judgments given when a target was correctly recalled ("Correct"), when a response was omitted ("Omission"), and when an incorrect response was given ("Commission"). I submitted the means (Figure 3.15) to a 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) x 3 (Recall Outcome: Correct vs. Omission vs. Commission) mixed-effects ANOVA with repeated measures on the last factor. Mauchley's Test is not

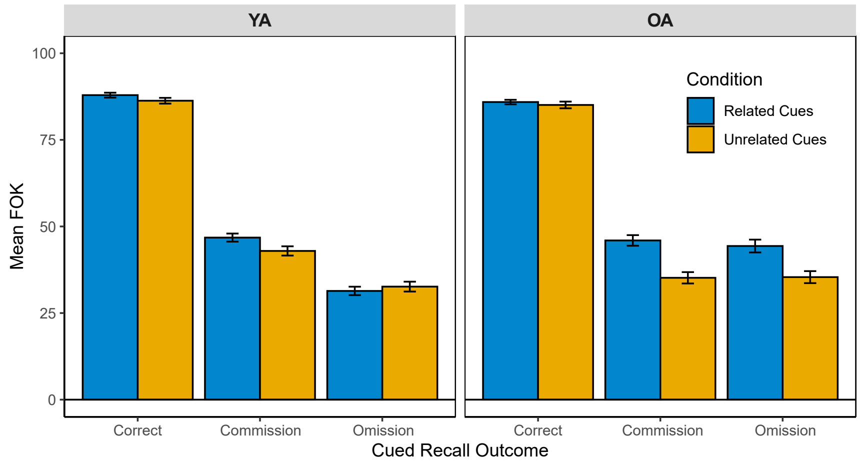

Figure 3.15: Average FOK judgments across recall outcomes, age groups, and experimental conditions.

significant, W = 0.99, p = 0.44, indicating that sphericity is not an issue in this analysis.

The 3-way interaction between Recall Outcome, Age Group, and Condition is significant, F(2,236) = 3.16, p = 0.04, η_p^2 = 0.01. An analysis of the means indicates that the effect is driven by differences between the experimental conditions for errors of omission and commission in both young and older adults. Specifically, young adults show significant differences across all recall outcomes, $M_{Correct}$ = 86.64, $SE_{Correct}$ = 2.35, M_{Comm} = 51.66, SE_{Comm} = 2.38, M_{Om} = 33.79, SE_{Om} = 2.54, but no significant differences between experimental conditions. Older adults, on the other hand, show significantly higher FOKs for correct items, M = 86.51, SE = 2.35, but no differences between errors of omission, M = 35.94, SE = 2.92, and commission, M = 50.85, SE = 2.70, although older adult participants in the Related Cues condition show greater mean FOKs for recall errors M_{Comm} = 58.05, SE_{Comm} = 3.79, M_{Om} = 39.89, SE_{Om} = 4.05, than those in the Unrelated Cues condition, M_{Comm} = 43.65, SE_{Comm} = 3.85, M_{Om} = 31.99, SE_{Om} = 4.20. A partial interaction contrast examining the mean differences in FOKs between the Related and Unrelated Cues conditions across recall errors for older adults is significant, F(1,411) = 4.41, p = 0.04, and indicates that the average difference in these FOKs is 13.89.

The main effect of Recall Outcome is also significant, F(2,236) = 321.13, $p < 0.01$, η_p^2

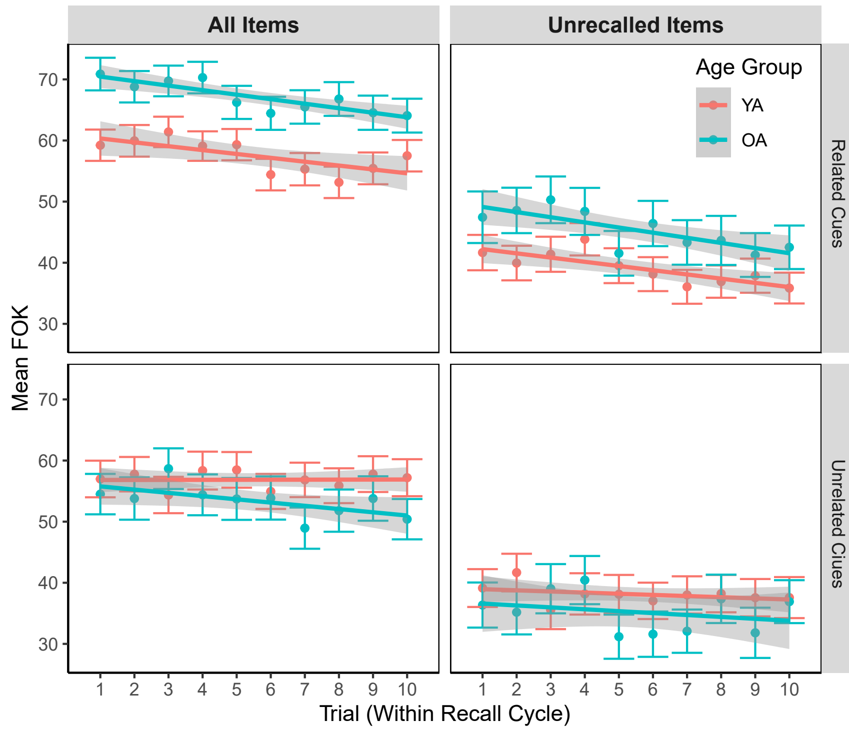

Figure 3.16: Average feeling-of-knowing judgments across trials, age groups, and experimental conditions for recalled and unrecalled items.

= 0.50, reflecting significantly higher mean FOKs for correctly recalled items, M = 86.75, SE = 1.77, compared to those for errors of commission, M = 51.25, SE = 1.80, $p <$ 0.01, and errors of omission, M = 34.87, SE = 1.94, $p <$ 0.01. Mean FOKs for errors of commission are also significantly higher than those for errors of omission, $p <$ 0.01. No other effects in this model are significant.

3.6.2 FOKs Across Trials

In the aggregate, average FOKs are expected to mirror recall performance, as FOKs are highly influenced by recall outcome (e.g. Krinsky & Nelson, 1985). Because participants in the Related Cues condition demonstrate the output interference effect, FOK judgments are expected to decrease across trials, similar to cued recall performance.

A larger question, however, is if FOK judgments show decreases over trials during output interference even when only examining judgments given to items that are not correctly recalled. Here, I provide analyses that examine the influence of output interference on mean FOKs for both all items and only for items that are not recalled.

All Items. Like recall, I examined mean FOKs for each trial (1 - 10), collapsed across all 4 blocks of recall (Figure 3.16). For participants in the Unrelated Cues condition, the 40 consecutive recall trials were separated into 4 consecutive blocks of 10 trials to provide a comparison to the Related Cues condition. These data were submitted to a GLM with Age Group (Young vs. Older Adults) and Condition (Unrelated vs. Related Cues) as categorical variables and Trial (1 - 10) as a continuous variable. Young adult participants in the Related Cues condition was set as the reference group.

Overall, the model provides an excellent fit to the data, R^2 = 0.92, F(7,32) = 49.23, $p < 0.01$. The Condition by Trial interaction is significant, F(1,32) = 4.51, p = 0.04, η_p^2 = 0.12, with a significant parameter estimate for participants in the Unrelated Cues condition, B = 0.64, t(32) = 2.27, p = 0.03. A simple slopes analysis indicates that mean FOKs significantly decrease over trials, but only for participants in the Related Cues condition, b = -0.69, t(32) = -4.84, $p < 0.01$. The same slope for participants in the Unrelated Cues condition is also negative, but only approaches significance, b = -0.26, t(32) = -1.84, p = 0.08.

The interaction between Age Group and Condition is also significant, F(1,32) = 129.72, $p < 0.01$, η_p^2 = 0.80, with a significant parameter estimate for older adult participants in the Unrelated Cues condition, B = -10.73, t(32) = -4.31, $p < 0.01$. A paired analysis using Tukey's method on the marginal means shows that the interaction is carried by significantly larger mean FOKs for older adult participants in the Related Cues condition, M = 67.13, SE = 0.58, compared to young adults in the same condition, M = 57.48, SE = 0.58, $p < 0.01$, to young adults in the Unrelated Cues condition, M = 56.86, SE = 0.58, $p < 0.01$, and to older adults in Unrelated Cues condition, M = 53.39, SE = 0.58, $p < 0.01$. The mean FOKs for older adult participants in the Unrelated Cues condition are significantly lower than those for young adults in the same condition, $p < 0.01$, as well as for young adults in the Unrelated Cues condition, $p < 0.01$. The difference in mean FOKs for young adults in the Related Cues and Unrelated Cues conditions is not significant, p = 0.87.

The main effect of trial within a recall cycle is significant, F(1,32) = 22.30, $p < 0.01$, η_p^2 = 0.41. The significant parameter estimate, B = -0.63, t(32) = -3.16, $p < 0.01$, indicates that, on average, participants report lower FOKs across recall trials.

Finally, the main effects of Condition, F(1,32) = 155.56, $p < 0.01$, η_p^2 = 0.83, and Age Group, F(1, 32) = 28.78, $p < 0.01$, η_p^2 = 0.47, are significant. An analysis of the marginal means indicates that these effects are due to higher FOKs in the Related Cues condition, M = 62.30, SE = 0.41, compared to the Unrelated Cues condition, M = 55.12, SE = 0.41, and to slightly higher FOKs for older adults, M = 60.26, SE = 0.41, compared to young adults, M = 57.17, SE = 0.41.

Unrecalled Items. I ran a second analysis using the same predictors, but with mean FOKs from unrecalled trials only as the outcome variable. Overall, the GLM accounts for a significant proportion of the variance in the data, R^2 = 0.79, F(7,32) = 17.37, $p < 0.01$. Unlike the previous model, however, the interaction between Condition and Trial only approaches significance, F(1,32) = 4.06, p = 0.05, η_p^2 = 0.11, and the parameter estimate for participants in the Unrelated Cues condition is not significant, B = 0.50, t(32) = 1.39, p = 0.18. Despite this, I examined the simple slopes for each of the conditions. Only the simple slope for the Related Cues condition is significantly different than zero, B = -0.77, t(32) = -4.23, $p <$ 0.01, and indicates that participants' FOKs for incorrect recall trials significantly decrease across trials when experiencing output interference. The same is not true for participants in the Unrelated Cues condition, whose slope is also negative, but is not significantly different than zero, B = -0.25, t(32) = -1.38, p = 0.18.

Like before, the interaction between Age Group and Condition is significant, F(1,32) = 38.66, $p < 0.01$, η_p^2 = 0.55, with a significant parameter estimate for older adult participants

		FOK x Recog Gamma (All)		FOK x Recog. Gamma (Unrec.)		FOK x Recall Gamma		FOK x CJ Gamma	
Age	Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
YA	Related	0.33	*0.03*	0.13	*0.04*	0.79	*0.04*	0.22	*0.03*
	Unrelated	0.35	*0.05*	0.14	*0.06*	0.81	*0.05*	0.25	*0.03*
OA	Related	0.30	*0.05*	-0.01	*0.08*	0.77	*0.04*	0.26	*0.03*
	Unrelated	0.23	*0.04*	0.05	*0.05*	0.81	*0.04*	0.26	*0.03*

Table 3.9: Average gamma correlation estimates for FOKs across age groups and conditions.

in the Unrelated Cues condition, B = -9.29, t(32) = -2.93, p = 0.01. A paired analysis using Tukey's method shows that the interaction is carried by a significantly larger mean FOK for older adult participants in the Related Cues condition, M = 45.34, SE = 0.73, compared to young adults in the Related Cues condition, M = 39.11, SE = 0.73, $p < 0.01$, to young adults in the Unrelated Cues condition, M = 38.10, SE = 0.73, $p < 0.01$, and to older adults in Unrelated Cues condition, M = 35.19, SE = 0.73, $p < 0.01$. The mean FOK for older adult participants in the Unrelated Cues condition is significantly lower than that for young adults in the same condition, $p < 0.01$, as well as for young adults in the Unrelated Cues condition, p = 0.04. The difference in mean FOKs for young adults in the Related Cues and Unrelated Cues conditions is not significant, p = 0.76.

Finally, both the main effects of Age Group, F(1,32) = 5.08, p = 0.03, η_p^2 = 0.14, and Condition, F(1,32) = 57.71, $p < 0.01$, η_p^2 = 0.64, are significant. The marginal means show that older adults reported larger FOKs on average for unrecalled items, M = 40.27, SE = 0.52, compared to young adults, M = 38.61, SE = 0.52, and that participants in the Related Cues condition reported larger FOKs, M = 42.23, SE = 0.52, than did participants in the Unrelated Cues condition, M = 36.65, SE = 0.52.

3.6.3 Relative Accuracy

A crucial question is whether FOK judgments under output interference accurately track future recognition memory outcomes and whether there are age-related differences in FOK accuracy under interference conditions. Here, I use canonical Goodman-Kruskal gamma correlations (L. A. Goodman & Kruskal, 1963; Krinsky & Nelson, 1985)—a non-parametric measure of association between two ordinal variables—to estimate the relationships between FOKs and other outcomes in this experiment, i.e. recognition accuracy, recall accuracy, and post-recognition confidence judgments (RCJs). Previous research has demonstrated that the gamma metric can be problematic (e.g. P. A. Higham and Higham, 2019; P. A. Higham et al., 2016; Masson and Rotello, 2009); however, these arguments are beyond the scope of the current study, and other analyses, such as a multilevel model using FOKs as the dependent variable, will help validate the gamma correlation estimates.

FOK x Recall Gammas. I first computed gamma correlations between FOK judgments and recall performance (Table 3.9) to estimate the degree to which individual cued recall outcomes affect FOKs, as previous research has shown that recall is a potent cue (e.g. Krinsky & Nelson, 1985). Overall, the gammas suggest that individuals relied heavily on experiences during cued recall when constructing FOKs, M = 0.79, SE = 0.02. Mean gammas for older adult participants in the Related Cues, M = 0.77, SE = 0.04, and Unrelated Cues, M = 0.81, SE = 0.04, conditions are high and virtually similar to those for young adults in the Related Cues, M = 0.79, SE = 0.04, and Unrelated Cues, M = 0.81, SE = 0.05, conditions. These data were submitted to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA. Like previous analyses, neither the interaction, F(1,146) = 0.06, p = 0.81, nor the main effects of age group, F(1,146) = 0.11, p = 0.74, and condition, F(1,146) = 0.12, p = 0.72, are significant. A Bayes Factor analysis indicated positive support in favor of the null hypothesis against the interaction, BF_{01} = 5.39, the main effect of age group, BF_{02} = 4.11, and the main effect of condition, BF_{03} = 3.40, further suggesting that there are no significant group differences.

FOK x Recognition Gammas (All Items). Next, I computed gamma correlations between each individual's FOK judgments and recognition outcomes (Table 3.9) to estimate how well these judgments track future memory outcomes—even in the presence of interference. Similar to previous FOK studies using similar stimuli (e.g. Eakin and Hertzog, 2012a; Hertzog et al., 2014), the mean gamma is above zero, M = 0.31, SE = 0.02, and indicates a small-to-moderate relationship between item-level FOKs and subsequent recognition memory outcomes. The mean gammas also indicate a slight numeric advantage in judgment accuracy for the young adult Related Cues, M = 0.33, SE = 0.03, and Unrelated Cues, M = 0.35, SE = 0.05, conditions compared to older adults in the Related Cues, M = 0.30, SE = 0.05, and Unrelated Cues, M = 0.23, SE = 0.04, conditions. A 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA on these data show no significant interaction between age groups and conditions, F(1,146) = 0.93, p 0.34, nor significant main effects of age group, F(1,146) = 0.29, p = 0.59, or condition, F(1,146) = 1.24, p = 0.27. A Bayesian Factor analysis indicates positive evidence in favor of the null (intercept) hypothesis (H_0), compared to the interaction between age group and condition, BF_{01} = 4.60, and the main effect of condition, BF_{02} = 5.08. The evidence in favor of the null model compared to a model with the main effect of age group is inconclusive, BF_{03} = 0.98. Given these results, the best interpretation of these data is that of no significant differences in mean gammas between age groups and experimental conditions, suggesting that FOK accuracy is similar across all participants.

FOK x Recognition Gammas (Unrecalled Items). Given the strong relationship between recall and FOKs, I additionally computed gamma correlations only for FOK and recognition trials in which the corresponding recall response was incorrect (Table 3.9). Discarding correct recall trials helps prevent FOK resolution measures from being anchored by judgments given during correct recall trials, as correctly-recalled items are extremely likely to be recognized later on and are consequently given FOKs close to ceiling (T. O. Nelson, 1984). Averaging these gammas across participants indicates that FOK resolution for unrecalled items is positive and slightly greater than zero, M = 0.08, SE = 0.03, indicating a small above-chance relationship. Further analyses indicate a slight numerical advantage for young adults in the Related Cues, M = 0.13, SE = 0.04, and Unrelated Cues, M = 0.14, SE = 0.05, conditions compared to older adults in the Related Cues, M = -0.01, SE = 0.08, and Unrelated Cues, M = 0.05, SE = 0.05, conditions, whose overall gammas are functionally zero. A 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) between-subjects ANOVA on these data also indicates non-significant results for the interaction between age group and condition, F(1,138) = 0.15, p = 0.69, the main effect of age group, F(1,138) = 2.55, p = 0.11, and the main effect of condition, F(1,138) = 0.40, p = 0.53. A Bayes Factor analysis indicates positive support for a null outcome for the interaction, BF_{01} = 4.96, the main effect of age group, BF_{02} = 1.28, and the main effect of condition, BF_{03} = 5.14, suggesting that these data most likely indicate no significant differences in FOK resolution for unrecalled items across age groups and experimental conditions, despite younger adults having numerically greater gamma correlations.

FOK x RCJ Gammas. Finally, I computed gamma correlations between FOKs and post-recognition retrospective confidence judgments (RCJs) to estimate the degree to which memory experiences during prospective memory judgments about recognition are shared with confidence judgments given after recognition memory trials (Table 3.9). Participants in this task showed a low-to-moderate correspondence between FOKs and RCJs, M = 0.25, SE = 0.01. Mean gammas for young adults in the Unrelated Cues condition, M = 0.25, SE = 0.03, older adults in the Related Cues condition, M = 0.26, SE = 0.03, and older adults in the Unrelated Cues condition, M = 0.26, SE = 0.03, were numerically higher than those for young adults in the Related Cues Condition, M = 0.22, SE = 0.03. These gammas were submitted to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Related Cues vs.

Unrelated Cues) between-subjects ANOVA. Like before, the interaction, $F(1,147) = 0.37$, $p = 0.54$, main effect of age group, $F(1,147) = 1.31$, $p = 0.25$, and main effect of condition, $F(1,147) = 0.63$, $p = 0.43$, did not reach significance. A Bayes Factor analysis indicates positive support for the interaction, $BF_{01} = 5.48$, main effect of age group, $BF_{02} = 4.94$, and main effect of condition, $BF_{03} = 5.45$; thus, the data support the hypothesis that young and older adults in the Related and Unrelated Cues conditions show the same degree of correspondence between FOKs made during recall and RCJs given after recognition.

3.6.4 Multilevel Model

In order to capture variations in FOK judgments within participants and blocked recall trials, I constructed a multilevel model using item-level FOKs as the dependent variable. This model not only helps validate the gamma correlations between FOKs and other variables, but it also allows these relationships to be modeled simultaneously. Of critical importance to this project are the relationships among FOKs, the trial within a recall block, and individual recall outcomes. For instance, if output interference occludes access to inaccessible memory items during recall, then the relationship between FOKs and recall trial (within a given cycle) is expected to be negative, but only for participants in the Related Cues condition. This negative slope should also be present even when item-level recall accuracy is accounted for.

Random Effects. The model assumes three levels of analysis: The level of the individual judgment, the level of the cycle within the recall sequence (i.e. 4 blocks of 10 recall trials), and the level of the participant engaging in the task. Thus, I specified two random intercepts: One at the level of the cycle within the recall sequence (despite it accounting for a small amount of residual variance; $ICC = 0.01$) and another at the level of the individual ($ICC = 0.36$). To determine whether these random variables are contributing a significant amount of information to the model, I used single-term deletion on both of the random intercepts and ran chi-squared tests on the changes in log-likelihood. Models that excluded the random intercept of recall cycle, $\chi(1)$ = 30.16, $p < 0.01$, and the random intercept of participant, $\chi(1)$ = 2158.36, $p < 0.01$, resulted in significant decreases in information compared to a model with both terms; thus, both random effects were retained in the final model.

Fixed Effects. The fixed effects from the multilevel model include the categorical variables age group (young vs. older adults), condition (Related Cues vs. Unrelated Cues), recall outcome (correct vs. incorrect), and recall cycle (1 - 4), along with the continuous predictor of recall trial within a cycle (1 - 10). Initially, all variables were entered and fully-crossed in the model, where the highest-order term was a 5-way interaction. For interpretability, as well as to prevent possible attenuation from non-significant higher-order interactions on lower-order interactions, I removed non-significant terms from the model until deletion resulted in a significant loss of information, starting with the higher-order interactons. As such, the 5-way interaction and all of the 4-way interactions were removed from the model without significant loss of fit compared to the full model, $\chi^2(16)$ = 14.11, p = 0.59. All lower-order interactions stayed in the model. Overall, the final trimmed model provided a good fit to the data: The pseudo-R^2 estimates for the total model and fixed effects only are 0.62 and 0.40, respectively, indicating that a large amount of information in the data is accounted for by the model.

Table 3.10: F-table for the multilevel model with item-level FOKs as the dependent variable.

	FOK	
	df	*F*
Age Group	1, 232.3	0.52
Condition	1, 231.3	**8.97***
Cycle	3, 5680.7	2.38+
Trial (Within Cycle)	1, 5678.6	**9.92***
Recall	1, 5733.5	**1414.26***
Age x Condition	1, 202.9	**4.55***
Age x Cycle	3, 5677.8	1.98
Age x Trial	1, 5678.4	0.93
Age x Recall	1, 5734.4	1.62
Condition x Cycle	3, 5687.2	0.86
Condition x Trial	1, 5678.5	**5.88***
Condition x Recall	1, 5733.6	**17.38***
Cycle x Trial	3, 5678.7	1.49
Cycle x Recall	3, 5686.3	0.58
Trial x Recall	1, 5684.3	0.47
Age x Condition x Cycle	3, 5674.3	0.84
Age x Condition x Trial	1, 5673.7	0.01
Age x Condition x Recall	1, 5805.9	**33.17***
Age x Cycle x Trial	3, 5674.0	0.67
Age x Cycle x Recall	3, 5688.5	0.65
Age x Trial x Recall	1, 5683.6	1.20
Condition x Cycle x Trial	3, 5674.3	1.82
Condition x Cycle x Recall	3, 5688.4	0.65
Condition x Trial x Recall	1, 5684.2	**4.84***
Cycle x Trial x Recall	3, 5684.2	1.20
Note:	$^{+}p<0.1$; $^{*}p<0.05$	

Table 3.11: Parameter estimates for the multilevel model using item-level FOKs as the dependent variable.

	Dependent Variable: FOK		
	B	*SE*	*p*
Intercept	42.96*	3.18	**<0.01**
Age Group (OA)	6.71	4.62	0.15
Condition (Unrelated)	1.46	4.48	0.75
Cycle 2	3.60	2.89	0.21
Cycle 3	1.37	2.96	0.64
Cycle 4	-4.56	2.92	0.12
Trial (Within Cycle)	-0.01	0.39	0.97
Recall (Correct)	43.46	2.89	**<0.01**
Age (OA) x Condition (Unrelated	-14.56	6.13	**0.02**
Condition (Unrelated) x Cycle 2	-5.76	3.57	0.76
Condition (Unrelated) x Trial	0.11	0.48	0.45
Condition (Unrelated) x Recall (Correct)	4.11	3.25	0.20
Cycle 2 x Trial	-0.56	3.23	0.37
Cycle 3 x Trial	-1.19	3.73	**<0.01**
Cycle 4 x Trial	-0.65	0.48	0.11
Cycle 2 x Recall (Correct)	-0.38	3.25	0.06
Cycle 3 x Recall (Correct)	-2.13	3.23	0.22
Cycle 4 x Recall (Correct)	3.00	3.27	0.81
Trial x Recall (Correct)	-0.06	0.40	0.21
Age (OA) x Condition (Unrelated) x Cycle 2	1.67	3.25	0.27
Age (OA) x Condition (Unrelated) x Cycle 3	5.03	3.23	**0.02**
Age (OA) x Condition (Unrelated) x Cycle 4	2.08	3.27	0.20
Age (OA) x Condition (Unrelated) x Trial	-0.04	0.40	0.92
Age (OA) x Condition (Unrelated) x Recall (Correct)	15.66	2.72	**<0.01**
Age (OA) x Cycle 2 x Trial	0.52	0.56	0.35
Age (OA) x Cycle 3 x Trial	-0.16	0.56	0.77
Age (OA) x Cycle 4 x Trial	0.41	0.56	0.47
Age (OA) x Cycle 2 x Recall (Correct)	1.44	3.31	0.66
Age (OA) x Cycle 3 x Recall (Correct)	0.96	3.31	0.77
Age (OA) x Cycle 4 x Recall (Correct)	-2.86	3.36	0.39
Age (OA) x Trial x Recall (Correct)	0.45	0.41	0.27
Condition (Unrelated) x Cycle 2 x Trial	0.22	0.56	0.69
Condition (Unrelated) x Cycle 3 x Trial	1.17	0.56	**0.04**
Condition (Unrelated) x Cycle 4 x Trial	0.79	0.56	0.16
Condition (Unrelated) x Cycle 2 x Recall (Correct)	-4.65	3.30	0.17
Condition (Unrelated) x Cycle 3 x Recall (Correct)	-2.94	3.31	0.38
Condition (Unrelated) x Cycle 4 x Recall (Correct)	-2.32	3.37	0.49

Condition (Unrelated) x Trial x Recall (Correct)	-0.90*	0.41	**0.03**
Cycle 2 x Trial x Recall (Correct)	0.21	0.57	0.71
Cycle 3 x Trial x Recall (Correct)	1.02	0.57	0.07
Cycle 4 x Trial x Recall (Correct)	0.48	0.58	0.41
σ^2 (Recall Cycle)	15.90		
τ^2 (Participants)	$<$0.01		
-2LL	52651.44		
R^2 (Fixed)	0.41		
R^2 (Overall)	0.62		

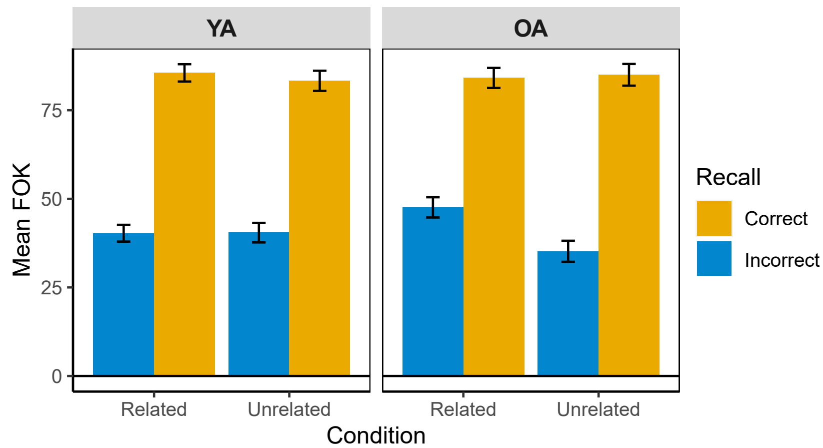

Figure 3.17: Marginal mean FOK judgments by age group, condition, and whether an item was recalled or not.

The 3-way interaction between condition, trial (within recall block), and recall accuracy is significant, F(1,5640.1) = 4.80, p = 0.03, with a significant slope estimate for participants in the Unrelated Cues condition for correctly-recalled items, B = -0.90, t = 5640.08, p = 0.03. However, neither of the simple slopes for participants in the Unrelated Cues condition when they correctly recall an item, b = -0.87, t = -2.05, p = 0.05, or do not recall the item at all, b = -0.11, t = -0.33, p = 0.74, are significantly different than zero, nor are the simple slopes for participants in the Related Cues condition who correctly recall an item, b = -0.06, t = -0.18, p = 0.86, or do not recall an item, b = -0.20, t = -0.59, p = 0.55. Interestingly, this indicates that the significance of the term is due to a significantly lower slope across trials for participants in the Unrelated Cues condition when target words are correctly recalled.

The 3-way interaction between age group, condition, and recall accuracy is also significant, F(1, 5805.9) = 33.17, $p < 0.01$, where the parameter estimate for older adults in the Unrelated Cues condition is significant for items that are correctly recalled, B = 15.66, t(5805.85) = 5.76, $p < 0.01$. An investigation of the marginal means (Figure 3.17) indicates that average FOK judgments for older adults in the Unrelated Cues condition who do not correctly recall an item, M = 35.20, SE = 2.97, are significantly lower than judgments given by older adults in the Related Cues condition who also do not correctly recall an item, M = 47.58, SE = 2.85, $p < 0.01$, and compared to judgments given by young adults who do not correctly recall a target word in both the Related Cues, M = 40.30, SE = 2.37, $p < 0.01$, and Unrelated Cues condition, M = 40.56, SE = 2.76. All mean FOKs for unrecalled items are significantly lower than the judgments given for young adults in the Related Cues condition, M = 85.54, SE = 2.45, young adults in the Unrelated Cues condition, M = 83.30, SE = 2.84, older adults in the Related Cues condition, M = 84.12, SE = 2.83, and older adults in the Unrelated Cues condition, M = 84.99, SE = 3.06, $ps <$ 0.01, when recall is successful. No other paired comparisons are significant.

The interaction between condition and recall accuracy is significant, F(1, 5733.6) = 17.38, $p < 0.01$, although the parameter estimate for participants in the Unrelated Cues condition who correctly recall an item is not significant, B = 4.11, t(5729.09) = 1.26, p = 0.21. Paired comparisons using Tukey's method indicate that FOKs given by participants in the Unrelated Cues condition during an unsuccessful recall attempt, M = 37.83, SE = 2.03, are significantly lower than those given by participants in the Related Cues condition during a successful recall attempt, M = 84.83, SE = 1.87, $p < 0.01$, and participants in the Unrelated Cues condition during a successful recall attempt, M = 84.15, SE = 2.09, $p < 0.01$, but not for FOKs given by participants in the Related Cues condition during an unsuccessful recall attempt, M = 43.94, SE = 1.85, p - 0.12, although FOKs for participants in the Related Cues condition given when recall fails is significantly lower than FOKs given when recall is successful, $ps < 0.01$. No other paired comparisons are significant.

The interaction between trial and condition is also significant, F(1, 5768.5) = 5.88, p = 0.02, although the parameter estimate for participants in the Unrelated Cues condition is not significant, B = 0.11, t(5678.55) = 0.23, p = 0.81. A simple slopes analysis also shows that neither the slopes for the Related Cues condition, b = -0.14, t(5678.55) = -0.54, p = 0.59, nor the slope for the Unrelated Cues condition, b = -0.43, t(5678.55) = -1.45, p = 0.15, are significantly different than zero.

The age group by condition interaction is significant, F(1, 202.9) = 4.55, $p < 0.01$, and while there is a significant parameter estimate for older adult participants in the Unrelated Cues condition, B = -14.56, t(253.48) = -2.38, p = 0.02, none of the paired comparisons between young adults in the Related Cues condition, M = 62.92, SE = 2.33, young adults in the Unrelated Cues condition, M = 61.88, SE = 2.72, older adults in the Related Cues condition, M = 65.85, SE = 2.75, and older adults in the Unrelated Cues condition, M = 60.10, SE = 2.93, were significantly different from each other, $ps > 0.05$.

The main effect of recall is significant, F(1, 5733.5) = 1414.26, and reflects the tendency for FOKs given during a successful recall attempt, M = 84.49, SE = 1.40, to be significantly greater than FOKs given during an unsuccessful recall attempt, M = 40.88, SE = 1.37, B = 43.46, t(5715.15) = 15.03, $p < 0.01$.

Additionally, the main effect of trial is significant, F(1, 5678.6) = 9.92, $p < 0.01$, although the parameter estimate itself is not, B = -0.01, t(5687.62) = -0.03, p = 0.97, indicating a numeric, but non-significant, decrease in FOK across trials.

Lastly, the effect of experimental condition is significant, F(1, 231.3) = 8.97, $p < 0.01$, but also with a non-significant parameter estimate for participants in the Unrelated Cues condition, B = 1.46, t(402.36) = 0.33, p = 0.75. Generally, participants in the Related Cues condition reported higher FOKs, M = 64.38, SE = 1.80, than participants in the Unrelated Cues condition, M = 60.99, SE = 2.00.

3.6.5 Summary

As expected, feelings-of-knowing are highly dependent upon cued recall performance. Participants in this task demonstrated this at both the aggregate level, where older adults in the Related Cues condition had significantly higher recall accuracy and average FOKs compared to all other groups, and at the item level, where gamma correlations between individual recall outcomes and FOK judgments were extremely high.

Importantly, FOKs appear to be influenced by output interference during cued recall.

While FOKs numerically decrease across trials for all participants, only the trial by FOK slopes are negative and significantly different than zero for participants who experience output interference during recall (i.e. Related Cues condition). This pattern holds even when isolating judgments that were given when recall was unsuccessful. The relationship between consecutive recall trials and FOKs is not clear-cut, however; the simple slopes remain negative, but do not reach significance when accounting for participant- and recall block-level differences in average FOKs in the multilevel context.

Relationships between FOKs and future recognition performance do not significantly differ between age groups or conditions, despite the influence of output interference on judgments. The gamma correlations for all items indicate a small-to-moderate relationship between FOKs and subsequent recognition performance while the gamma correlations for judgments given when recall is unsuccessful are low for young adults and virtually zero for older adults (Table 3.1).

CHAPTER 4
GENERAL DISCUSSION

In the study reviewed here, output interference during recall was expected to affect memory and FOK accuracy. Instead, judgment accuracy was not different between participants that studied word pairs in which the cue words were related (leading to output interference) and those that studied completely unrelated cue-target pairs, nor were there differences between young and older adults. These null effects are present despite the fact that participants in the Related Cues condition showed significant decreases in cued recall accuracy over successive trials, as well as the fact that older adult participants in the Related Cues condition recalled significantly more target words on average than all other participants. The results reviewed here provide continued support for similarities in episodic FOK accuracy across the lifespan (Eakin et al., 2014; Hertzog, Dunlosky, & Sinclair, 2010; Hertzog et al., 2014; MacLaverty & Hertzog, 2009), particularly under interference conditions (Eakin & Hertzog, 2006, 2012a, 2012b).

Importantly, the null effects at the aggregate level are not indicative of patterns in responses at the trial-level. Participants that studied word pairs with related cue words demonstrated significant decreases in cued recall performance across blocked trials, along with increases in errors of omission, compared to participants who studied completely unrelated cue-target pairs, where recall performance and error rates were stable across trials. Accordingly, self-reported recollection during cued recall significantly decreased across trials while states of unknowing (i.e. "No Memory") increased. Finally, participants that experienced output interference during recall exhibited significant decreases in their FOK judgments across trials.

The first section of this manuscript introduced two accounts of output interference based on the extant literature: 1) The *competition hypothesis*, where OI is hypothesized to be a function of co-activated items; and 2) the *retrieval-suppression hypothesis*, where OI is hypothesized to be a function of previously-activated items occluding retrieval of a target item. Based on the results of this study, I argue that these data are more consistent with a retrieval-suppression account of output interference. The remaining sections of this chapter review the results of the study in more detail and weigh evidence for and against this hypothesis.

4.1 Output Interference, Cued Recall, and Aging

4.1.1 Recall Accuracy

Research on output interference spanning almost 60 years (Tulving & Arbuckle, 1963; Wilson et al., 2020) indicates that shared retrieval cues induce decreases in cued recall accuracy across successive trials. Participants in the current study exhibited a similar pattern: For those in the Related Cues condition, cued recall accuracy decreased by an average of 1.8% for each successive trial within a block of trials, while changes in recall accuracy across trials for those in the Unrelated Cues condition were positive, but not significant. Despite wide-spread empirical evidence of age-related deficits in LTM (Kausler, 1994), such as a reduced ability to inhibit irrelevant information during retrieval (Lustig et al., 2001), young and older adult participants exhibited similar patterns of OI during cued recall. Age invariance in OI during cued recall is a novel finding, but is consistent with the findings of Smith (1975), where young, middle, and older adults showed similar decreases across consecutive responses in a free recall task.

A notable departure in the results of the current study compared to earlier studies is not only the relative equivalence in cued recall performance between young and older adults, but also the advantage in recall accuracy for older adults in the Related Cues condition, whose recall performance was significantly higher than older adults in the Unrelated Cues condition and young adults in both experimental conditions. The simplest explanation for this finding is that mean recall performance is being anchored by 1 or 2 high-performing outliers; however, the data indicate that 11 of the 35 older adult participants in the Related Cues condition had cued recall accuracy averages above 80%, making the possibility of a couple of outliers improbable. Another potential explanation for this finding is the fact that some older adults in this study presented higher baseline cognition than older adults in other cross-sectional studies. Indeed, participants recruited from Amazon's Mechanical Turk provide high-quality research data (J. K. Goodman et al., 2013), but are often better-educated than a similar normative sample of adults in the United States (Levay et al., 2016). All of the older adults recruited for this study had at least a 4-year college education, and several had master's or professional degrees. Another explanation for this result could be that older adults felt more compelled to devote cognitive resources to a task that has the potential to evoke aging stereotypes about memory (Hess, 2014; Hess et al., 2003). These explanations are overshadowed by the fact that performance was not equivalent across experimental conditions for older adults; in fact, older adults in the Unrelated Cues condition performed similarly to young adults in both the Related and Unrelated Cues conditions. A final and most likely explanation, then, is that this difference in cued recall performance is an artifact of sampling and would likely disappear with a larger sample size.

4.1.2 Memory Errors

Of equal importance to recall accuracy are the patterns of errors that participants exhibit during cued recall. In the current study, decreases in recall accuracy across trials were mirrored by increases in retrieval failures, or errors of omission, but only for participants who studied word pairs with semantically-related cues. The probability of providing an incorrect response during recall, or an error of commission, did not change across trials for participants in the OI condition, and the separate probabilities of correct recall, errors of omission, and errors of commission also did not change across trials for participants who did not experience OI. This is consistent with previous studies examining errors during cued recall during OI (Cox et al., 2018; Wilson & Criss, 2017; Wilson et al., 2020). These error patterns were the same for both young and older adult participants.

4.1.3 Heterogeneity in OI

The multilevel model analyses, which help control for block- and participant-level variation in recall performance, show that accuracy and error rates across trials are different across blocks of recall trials for both young and older adults. Heterogeneity in the OI effect certainly complicates the interpretation of the results in this study, but has precedent in the extant literature. Roediger (1973), for example, reported significant differences in recall patterns across successive blocks in a task in which participants were asked to freely recall items belonging to a category. In a later review paper, Roediger (1974) highlights several instances in which differences in recall performance due to OI were best examined using alternative methods, such as comparing performance in the first half of the trials to the second half (e.g. Kay & Poulton, 1951), or testing linear decreases across all trials, but only for grouped lists with 6 or fewer words (e.g. Arbuckle, 1967). Recent OI studies do not examine memory accuracy and error trends in such detail, but instead elect to use Bayesian analyses that contrast a general hypothesis of changes across trials to a null hypothesis of no changes (Wilson et al., 2020).

4.2 Metamemory and Output Interference

4.2.1 Overall FOK Magnitude and Accuracy

Contrary to the expected results, output interference during cued recall did not affect metamemory accuracy: The relationship between FOKs and subsequent recognition performance for all participants was small-to-moderate, M = 0.31, SE = 0.02; however, when FOKs are limited to trials in which cued recall was unsuccessful, this relationship is functionally zero, i.e. near-chance, M = 0.08, SE = 0.03. Even though FOKs are highly correlated with recall for both experimental conditions, and despite the fact that recall accuracy decreased across trials for participants in the Unrelated Cues condition, OI did not significantly alter FOK accuracy in this task. Furthermore, gamma correlations between FOKs and RCJs are low-to-moderate for all participants, M = 0.25, SE = 0.01, and suggests that there are no group differences with respect to the post-retrieval experiences that are shared between cued recall and 4AFC tests. Age- and condition-invariance in the current study is consistent with previous FOK research. Eakin and Hertzog (2012a), for example, found that FOK resolution was above-chance only when examining all items compared to only unrecalled items, with no significant interactions or main effects. The data from the current study affirm the extant finding that FOK accuracy is equivalent between young and older adults in interference paradigms.

The main influences on FOKs in this study were cued recall outcomes, as shown by high FOK by recall gamma correlation across all participants, M = 0.79, SE = 0.02. This is consistent with the results of most other FOK studies (c.f. Krinsky & Nelson, 1985), where the highest FOKs are typically given to correctly-recalled items, FOKs near the middle of the range are given to errors of commission, and the lowest FOKs are given to errors of omission. Interestingly, mean FOKs for each of the different recall errors were consistent across age groups and conditions: Mean FOKs when cued recall is correct, M = 86.57, SE = 1.770, are significantly higher than those for errors of commission, M = 51.25, SE = 1.77, $p < 0.01$, which in turn are significantly greater than those for errors of omission, M = 34.87, SE = 1.94, $p < 0.01$. The only deviation from this pattern of findings was found in older adults, where participants in the Related Cues condition had higher mean FOKs for errors of omission, M = 44.35, SE = 1.85, and commission, M = 45.96, SE = 1.53, compared to mean FOKs for errors of omission, M = 35.37, SE = 1.73, and commission, M = 35.18, SE = 1.64, for participants in the Unrelated Cues condition. Significant differences in cued recall performance in older adult learners between the two conditions, however, suggest that these differences in mean FOKs between error types are merely due to higher memory performance for older adults in the Related Cues condition.

These results are inconsistent with the findings of previous metamemory studies. Schreiber and colleagues (Schreiber, 1998; Schreiber & Nelson, 1998), for instance, demonstrate that FOK magnitude is inversely related to the amount of competition present during memory retrieval. Maki (1999) shows a similar pattern of results, where overall FOKs are lower for items that were repeated (i.e. retroactive interference)—a pattern of results that was replicated for iJOLs by Diaz and Benjamin (2011). Eakin (2005) reports that FOKs and predictions made prior to retrieval (i.e. predictions of knowing; POK) are higher on average for items under retroactive interference. Eakin and Hertzog (2006, 2012a, 2012b) replicate this pattern of results and demonstrate that both young and older adults provide significantly higher POKs, FOKs, and dJOLs under implicit interference conditions. Each of these studies conclude that aggregate metamemory judgments are sensitive to interference during retrieval; however, this is not the case in the present study, where group-level FOKs are not different between OI conditions, even when examining judgments by recall outcome. In the aggregate, both young and older adults do not appear to be sensitive to the effects of OI during cued recall when reporting FOKs.

4.2.2 Metacognitive Judgments Over Trials

While metamemory judgments do not show sensitivity to OI in the aggregate, they do at the trial-level. Both young and older adults' FOKs decrease across trials (within a recall block) during OI at retrieval. This pattern emerges even after accounting for variation at the level of the individual and recall block, as well as when examining FOKs for unrecalled items only. The negative slope across trials for participants that experience OI remains significant even when accounting for significant decreases in all FOKs across trials (e.g. Dunlosky & Matvey, 2001). Thus, both young and older adults' FOKs accurately tracked item-level recall accuracy across trials.

A novel finding in this study is the successive decrease in FOK magnitude across trials during interference. The negative slope is based on item-level responses, meaning that participants in the Related Cues condition generally provide lower FOKs for each recall trial.

This is curious for a few different reasons. First, these decreases are inconsistent with the competition account of FOK construction described by Schreiber (1998) and Schreiber and Nelson (1998). The authors argue that decreases in FOK magnitude are directly related to the amount of competition between neighboring concepts by the test cue, as demonstrated through manipulations of cue set size. In this study, however, the relationships between cue words and their superordinate categories do not change across trials, meaning that changes in judgments in this task cannot be attributed to cue set size across trials. Instead, the changes in FOKs in this task represent changes across trials that are reflected in the retrieval process itself rather than the relationships between a cue word and neighboring concepts (M. C. Anderson & Neely, 1996). Second, participants that experience OI report successively decreasing FOKs even when only examining unrecalled items. Even though average FOKs are larger for errors of commission than for errors of omission, these changes in FOKs across trials for unrecalled items cannot be due to changes in rates of errors of commission, as the probability of providing an incorrect answer during cued recall does not significantly change across trials. The changes in FOKs for incorrect recall trials reflect the changes in trace access to items in memory across successive trials, even when participants are unable to provide a response during recall.

Patterns in “Remember”, “Know”, and “No Memory” judgments in this task also reflect changes in recall performance: For both young and older adult participants that experience OI, self-reported recollection decreases across trials while states of “No Memory” increase across trials, which are consistent with the trade-off between decreases in cued recall accuracy and errors of omission. Overall, both young and older adults show similar patterns of recollection during retrieval (MacLaverty & Hertzog, 2009). Interestingly, reported states of familiarity decrease across trials for young adults, but increase across trials for older adults. The reason for this discrepancy is unclear, however; previous studies exploring R/K/N judgments and aging have found similar patterns, where decreased states of recollection are accompanied by increased familiarity for older adults (Daniels et al., 2009;

Koen & Yonelinas, 2016; Perfect & Dasgupta, 1997), but typically with significant differences in recollection between age groups. Conversely, in the current study, young and older adults had similar levels of endorsement of recollection and familiarity during cued recall, along with similar recall outcomes. Given that young and older adults show similar rates of correct recall, omission, and commission, the most likely explanation is that older adults interpret the experience of OI differently than young adults, even though memory performance is the same for both. If such a difference in interpretation exists, then it is localized to states of familiarity and not to any other metacognitive state.

4.3 Theoretical Accounts of Output Interference

The results of the current study are consistent with current mechanistic accounts of OI. During recall, the process of successfully retrieving an item strengthens the representation of that item (Carrier & Pashler, 1992; Roediger & Karpicke, 2006), sometimes referred to as "incrementing" in computational models of memory (Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997). Additionally, memory search during retrieval is limited in capacity, where an individual can attempt only a certain number of retrieval attempts during a given trial (i.e. a "retrieval filter"; Dougherty & Hunter, 2003a, 2003b; Rundus, 1973; R. P. Thomas et al., 2008). During output interference, successful retrieval of a word pair early in a recall sequence strengthens its representation, which leads to the same word pair being repeatedly sampled in later trials when cue words share semantic information. Oversampling of a previously recalled item in later trials causes the retrieval filter to reach its maximum number of attempts, reducing the probability of the correct target being sampled. Increased limitations in sampling during retrieval, in turn, result in increases in retrieval failures during cued recall, i.e. errors of omission (Wilson et al., 2020).

Most accounts of interference during memory favor theoretical accounts that espouse the influence of competing items during retrieval, i.e. a competition account. This general claim is supported by most research on memory interference, where experimental manipulations of the stimuli directly increase the number of competitors during a memory task (e.g., extra-list cueing; Eakin & Hertzog, 2012a), leading to decreases in accuracy and increases in false memory. The strength of an interference effect in these experiments is directly linked to WM, where decreases in executive processes lead to failures to inhibit highly-activated lures (e.g. M. C. Anderson & Levy, 2007). In OI, however, the stimuli used in the experiment are controlled on the basis of semantic properties in order to eliminate differences in spreading activation. Further, OI during retrieval does not lead to increases in familiarity and errors of commission during recall, as would be expected from a competition account; rather, OI leads to increases in errors of omission and "No Memory" states. Finally, young and older adults show functionally similar recall and judgment patterns when confronted with OI, with the exception of self-reported familiarity.

Given the increases in retrieval errors and feelings of no memory shown in this study, a competition account of OI is unlikely. A more likely explanation is one that implicates interference as a product of the retrieval mechanism itself, wherein over-sampling of a previously-recalled item with a related cue precludes retrieval of the correct target memory. Increases in errors of omission and feelings of no memory indicate that participants are unlikely to access information about the target memory during OI, where the previously-activated item precludes target items from sampling. The patterns memory and metamemory judgments in this task, then, are more consistent with a retrieval-suppression account of OI as opposed to a competition account.

While the results are consistent with suppression of a target item during OI, it is possible that they merely have the appearance of being the result of occlusion. A final possibility—one that was not specified prior to this study—is a misattribution account of OI, where learners are able to access information to a target item during OI, but falsely attribute it to a previously-recalled item. Consider a situation in which a learner is attempting to recall a target word during an OI study. In later trials, the cue word may induce activation of information related to the correct target word, or even the target word itself; however, the heightened activation of a previously-recalled word pair with a shared cue may overshadow these newer activations and lead the learner to falsely attribute the newly-activated information from the true cue-target pair to the highly-activated cue-target pair that was recalled earlier. Even though the learner may be unable to correctly attribute the new activations to the correct target word, she would still be able to leverage this non-criterial recollection for an FOK, leading to higher FOK resolution than expected under occlusion. This account could plausibly explain both the OI effects during recall and the null effects in FOK resolution, although identifying errors stemming from misattribution is not possible in these data.

4.4 Conclusions

The results of the current study are consistent with a growing body of research finding no age differences in FOK accuracy (e.g., Hertzog, Dunlosky, & Sinclair, 2010; Hertzog et al., 2014; Hertzog et al., 2002), but are inconsistent with respect to previous interference studies that examine FOKs (Eakin & Hertzog, 2006, 2012a). Unlike previous studies, aggregate memory performance and metamemory judgments were no different between age or interference conditions; however, differences do exist at the trial-level, where FOKs and reports of recollection decrease across trials and errors of omission and states of unknowing increase for participants experiencing OI. Taken together, the specific changes across trials indicate that participants of all ages are generally aware of the effects of OI (that is, a gradual decrease in recall accuracy across trials) and are providing memory responses and judgments that implicate suppression, and not competition, as the mechanism underlying interference during memory retrieval.

4.4.1 Future Directions

A line of study that would strengthen mechanistic accounts of the OI effect is one that is able to manipulate the strength of interference in these tasks. In previous retrieval-induced forgetting (RIF) research, for instance, the degree to which a response was inhibited was dependent upon the normative frequency of an item in a category (M. C. Anderson et al., 1994). In studies of implicit interference, the number of associates co-activated with a given cue word (i.e. cue-set-size) is directly linked to memory and metamemory performance (Eakin & Hertzog, 2006, 2012a; D. L. Nelson & McEvoy, 1979; Schreiber, 1998; Schreiber & Nelson, 1998). For OI, it is possible that the effect is most dependent upon the strength of the cue with respect to its taxonomic category, but only for the first few items during a retrieval task. Highly typical cue words are likely to evoke a strong activation during retrieval and would likely be re-sampled at higher rates in later trials than cue words of lower typicality if presented early in a retrieval task. Earlier OI research has demonstrated that the strength of the effect is not dependent upon the size of a category (Roediger & Schmidt, 1980), although more recent research has challenged this view (Bäuml, 1998). To my knowledge, however, a direct link between semantic attributes of a cue and its position in a retrieval sequence has yet to be examined.

Another helpful line of study would be one that examines retrieval latency and its relationship to responses during OI[1]. Despite reaction times being an integral part of memory research (e.g., J. R. Anderson, 1974), only a few studies have reported response durations under conditions similar to output interference. Bousfield and Sedgewick (1944), for example, demonstrate that responses are quick at first, but decrease with time when participants are asked to name exemplars of well-learned categories, such as types of birds. More recently, Rohrer and Wixted (1994), provide evidence that response latency is inversely related to list length in free recall (also see Wixted & Rohrer, 1994) through analyses of reaction times fit to ex-Gaussian distributions. A direct examination of RTs in OI tasks would help connect memory outcomes and experiences to proposed mechanisms of interference.

Finally, the OI effect presents a unique opportunity for examining the influences of executive attention, inhibition, and aging on the retrieval process. Several theoretical ac-

[1]Such an analysis would have been conducted in the current study, however, recall latencies were not recorded due to a coding error.

counts of memory retrieval implicate a retrieval filter, or a maximum number of retrieval attempts before search is terminated (Raaijmakers & Shiffrin, 1981; Rundus, 1973; Shiffrin & Steyvers, 1997), in OI effects (Criss et al., 2011; Koop et al., 2015; Wilson et al., 2020). Further, several studies have indicated that the comparison process during memory retrieval is limited by working memory capacity (Dougherty & Hunter, 2003a, 2003b; R. P. Thomas et al., 2014; R. P. Thomas et al., 2008), a hypothesis that is partly supported by relationships between WM and interference effects in young and older adults (M. C. Anderson & Levy, 2007; Aslan & Bäuml, 2011; Lustig et al., 2001; Williams & Zacks, 2001). The experimental paradigm used in the current study can be used to investigate capacity limitations (i.e. the number of retrieval attempts) due to executive deficits during retrieval. For example, under a hypothesis that young adults have greater WM capacity and, in turn, fewer capacity limitations during retrieval, young adults would be expected to show little or no signs of OI for shorter paired associates lists with related cue words compared to older adults. Assuming that OI is due the effects of over-sampling a highly-activated item, such a result would indicate that older adults entertain significantly fewer potential responses during memory retrieval compared to young adults, causing them to reach their maximum number of retrieval attempts without successfully sampling the correct target word, even when there are relatively few items to sample from. Such a comparison is not possible in the current study, where all participants studied the same number of paired associates in each block. Even so, both young and older adults showed similar OI effects, despite older adults scoring significantly lower on the Pattern Comparison Task—a common measure of processing speed in psychological research (Salthouse, 1992, 1996).

Appendices

APPENDIX A
ADDITIONAL ANALYSES

A.1 Curvilinearity in Recall Accuracy Across Trials

Similar to the cued recall analyses in Section 3.2.2, I averaged recall performance across the 4 blocks of cued recall trials for both age groups in both the Related Cues and Unrelated Cues conditions. Afterwards, I tested for curvilinearity in cued recall performance across trials within blocks by first centering the trial numbers and then exponentially transforming these numbers into quadratic and cubic terms[1] and submitting these data to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Unrelated Cues vs. Related Cues) x 10 (Trial Within Recall Cycle - Linear: 1 - 10) x 10 (Trial Within Recall Cycle - Quadratic: 1 - 10) repeated-measures GLM.

Like before, only the Age Group by Condition interaction is significant, $F(1,1482) = 34.42$, $p < 0.01$, $\eta^2 = 0.02$; however, the main effects of Condition, $F(1,1482) = 16.50$, $p < 0.01$, $\eta^2 = 0.01$, and Age Group, $F(1,1482) = 24.57$, $p < 0.01$, $\eta^2 = 0.02$, are also significant. The main effects of linear, $F(1,1482) = 1.42$, $p = 0.23$, $\eta^2 < 0.01$, quadratic, $F(1,1482) = 0.01$, $p = 0.97$, $\eta^2 < 0.01$, and cubic, $F(1,1482) = 0.16$, $p = 0.69$, $\eta^2 < 0.01$, centered trial number are not significant. Thus, there is evidence for neither a curvilinear nor a linear decrease in recall across trials.

A.2 Relationships Between 4AFC and Recognition Trials

Unlike cued recall, trials during recognition memory were completely randomized in this task. Even so, it is possible that output interference builds across recognition trials (Criss et al., 2011), even when trials with related stimuli (i.e. exemplars from the same

[1] These terms were limited to the second and third orders for ease of model estimation and interpretation.

Table A.1: Logistic model predicting recognition outcome (in log-odds) from age group, experimental condition, and trial within category during recognition.

	Dependent variable:
	Recognition
Intercept	**0.805***
	(0.105)
Age (OA)	−0.015
	(0.164)
Condition (Unrelated)	−0.208
	(0.159)
Trial (Within Category)	−0.016
	(0.017)
Age (OA) x Condition (Unrelated)	0.004
	(0.240)
Age (OA) x Trial	0.035
	(0.026)
Condition (Unrelated) x Trial	0.020
	(0.026)
Age (OA) x Condition (Unrelated) x Trial	−0.034
	(0.039)
-2LL	7634.13
AIC	7,650.133
Δchi^2	17.14*
Note:	$^+$p<0.1; *p<0.05

category) are not blocked. In order to analyze this, I renumbered the recognition trials for participants in the Related Cues condition so that they represent the order in which each set of stimuli (out of 10) within a given category (out of 4) was presented. For participants in the Unrelated Cues condition, I dummy-coded trials within blocks by numbering the 40 consecutive trials 1 through 10 over 4 "pseudo" blocks. These pseudo-blocks of consecutive trials were used to compare recognition performance across trials within blocks for participants in the Related Cues condition.

I submitted the binary recognition data (i.e. correct recognition vs. incorrect recognition) to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Unrelated Cues vs. Related Cues) x 10 (Trial Within Category: 1 - 10) GLM with a logit link on the dependent

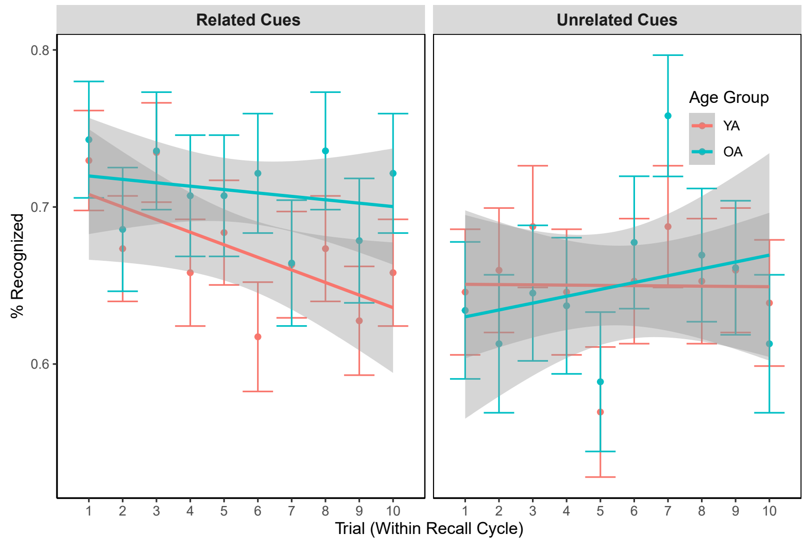

Figure A.1: Average 4-alternative forced choice recognition memory performance by age group and cue-relatedness across consecutive cued recall trials from earlier in the task.

variable (Table A.1). The model accounts for a significant amount of information compared to a null (intercept) model, $\chi^2(7) = 17.14$, $p = 0.02$; however, none of the parameter estimates are significant, with the exception of the intercept, which indicates that participants are 2.24 times more likely to correctly recognize an item during 4AFC than not, Exp(B) = 2.24, $z = 7.69$, $p < 0.01$. The model affirms the non-significant age group and condition effects from earlier analyses, but also indicates that the odds of correctly recognizing an item does not change across trials.

A.3 Relationships Between 4AFC and Earlier Recall Trials

A similar question is whether output interference during recall directly affects the probability of correctly recognizing an item later on. To examine this, I performed a similar analysis to that conducted during the cued recall analyses (section 3.2) by back-sorting item-level recognition outcomes to their corresponding recall trials (within block) earlier in the task. If output interference during recall leads to similar memory performance during recognition, then there should be significantly lower trial by memory performance slopes for participants in the Related Cues condition compared to the Unrelated Cues condition,

Table A.2: Logistic model predicting recognition memory outcome (in log-odds) from age group, experimental condition, and trial within recall blocks.

	Dependent variable:
	Recog_ACC
Intercept	**0.883***
	(0.091)
Age (OA)	0.060
	(0.143)
Condition (Unrelated)	-0.260^{+}
	(0.137)
Trial (Within Cycle)	$-$**0.036***
	(0.017)
Age (OA) x Condition (Unrelated)	-0.151
	(0.208)
Age (OA) x Trial	0.026
	(0.027)
Condition (Unrelated) x Trial	0.036
	(0.026)
Age (OA) x Condition (Unrelated) x Trial	-0.006
	(0.039)
Observations	6,039
-2LL	7630.12
Δchi^2	21.06*
Note:	$^{+}$p<0.1; **p<0.05

similar to cued recall performance.

I submitted the binary recognition outcomes to a 2 (Age Group: Young vs. Older Adults) x 2 (Condition: Unrelated Cues vs. Related Cues) x 10 (Trial Within Recall Cycle: 1 - 10) GLM with a logit link on the dependent variable (Table A.2). A comparison to the null (intercept) model indicates that the full model accounts for a significant amount of information in the data, χ^2(7) = 21.06, $p < 0.01$.

Only the parameter estimate for trial is significant, Exp(B) = 0.96, z = -2.17, p = 0.03, and indicates that the odds of providing an incorrect recognition response decreases by 1.04 for each consecutive trial within a block during recall. This indicates that changes recognition performance are connected to serial position during cued recall, although this effect is weak.

A.4 ANOVA - R/K/N Judgments

I submitted the averages of each of the three ratings (R/K/N) for each participant were submitted to a 3 (Rating: *R* vs. *K* vs. *N*) x 2 (Age Group: Young Adults vs. Older Adults) x 2 (Condition: Related Cues vs. Unrelated Cues) mixed-effects ANOVA, with repeated measures on the first factor (Figure 3.10). Mauchly's Test for Sphericity is significant on the repeated measure, W = 0.897, $p < 0.01$; accordingly, I adjusted the degrees of freedom for all F-tests that include the Rating factor using the Hyunh-Feldt correction.

The three-way interaction between Rating, Condition, and Age Group is significant, F(1.8, 269.8) = 3.13, p = 0.04, η_p^2 = 0.02, and is qualified by a two-way interaction between Rating and Condition, F(1.8, 269.8) = 3.549, p = 0.03, η_p^2 = 0.02. I decomposed these interactions by running 2 separate 3 (Rating: *R* vs. *K* vs. *N*) x 2 (Condition: Related Cues vs. Unrelated Cues) mixed effects ANOVAs for both of the age groups (i.e. YAs and OAs). For young adults, whose data violated the assumption of sphericity, W = 0.746, $p < 0.01$, only the main effect of Rating is significant, F(1.6,134.5) = 4.84, p = 0.01, η_p^2 = 0.06. Pairwise t-tests indicate that young adults reported "Know" significantly fewer times on average, *M* = 0.25, *SE* = 0.02, compared to both "Remember", *M* = 0.33, *SE* = 0.02, p = 0.02, and "No Memory" (*M* = 0.38, *SE* = 0.03, $p < 0.01$). For older adults, only the Rating by Condition interaction is significant, F(2.0, 129.2) = 5.27, $p < 0.01$, η_p^2 = 0.08. Separate one-way ANOVAs indicate no significant differences between average rates for R/K/N judgments for older adults in the Related Cues condition, F(2.1, 72.0) = 0.94, p = 0.40, but significant differences for older adults in the Unrelated Cues condition, F(1.6, 49.2) = 7.29, $p < 0.01$, η_p^2 = 0.20. Pairwise t-tests indicate that older adults in the Unrelated Cues condition reported significantly higher rates of "No Memory" judgments, M = 0.25, SE = 0.04, than "Know", M = 0.18, SE = 0.03, p = 0.01, or "Remember", M = 0.17, SE = 0.03, p = 0.03, judgments. A Bayes Factor analysis on the full (3x2x2) ANOVA suggests strong evidence in support of the null hypothesis of a main effect of Rating compared to the main effect of Age Group (BF_{01} = 338.27); main effect of Condition (BF_{02} = 364.06); interaction between Age Group and Condition (BF_{03} = 361.11); interaction between Age Group and Rating (BF_{04} = 386, 65); interaction between Condition and Rating, (BF_{05} = 57.28); and the three-way interaction between Age Group, Condition, and Rating (BF_{05} = 16.65).

The interaction between Condition and Rating is also significant, F(1.8, 269.8) = 3.55, p = 0.03, η_p^2 = 0.03. The effect is driven by significantly higher rates of *R* for participants in the Related Cues condition, M = 0.35, SE = 0.03, compared to *K* rates for participants in the Unrelated Cues condition, M = 0.24, SE = 0.03, p = 0.04; significantly higher rates of *N*, M = 0.42, SE = 0.03, compared to *R*, M = 0.30, SE = 0.03, p = 0.03, for participants in the Unrelated Cues condition; and significantly higher rates of *N* compared to *K*, M = 0.24, SE = 0.03, $p < 0.01$, for participants in the Unrelated Cues condition. *K*, M = 0.29, SE = 0.03, and *N*, M = 0.32, SE = 0.03, rates do not significantly differ from any other rating across conditions, $ps > 0.05$.

Finally, the main effect of Rating is significant, F(1.84, 269.8) = 5.55, $p < 0.01$, η_p^2 = 0.04, reflecting significantly lower rates of "K", M = 0.27, SE = 0.02, compared to *N*, M = 0.37, SE = 0.02. The difference between endorsement rates for *K* and *R*, M = 0.33, SE = 0.03, approaches, but does not reach significance, p = 0.06. No other paired comparisons or effects in this model are significant.

REFERENCES

Allen-Burge, R., & Storandt, M. (2000). Age equivalence in feeling-of-knowing experiences. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *55*(4), P214–P223.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review*, *7*(3), 522–530.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087.

Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 81–102). American Psychological Association.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 237–313). Elsevier.

Arbuckle, T. Y. (1967). Differential retention of individual paired associates within an RTT ”learning” trial. *Journal of Experimental Psychology*, *74*(3), 443–451.

Aslan, A., Bäuml, K.-H., & Pastötter, B. (2007). No inhibitory deficit in older adults' episodic memory. *Psychological Science*, *18*(1), 72–78.

Aslan, A., & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 264–269.

Badham, S. P., Poirier, M., Gandhi, N., Hadjivassiliou, A., & Maylor, E. A. (2016). Aging and memory as discrimination: Influences of encoding specificity, cue overload, and prior knowledge. *Psychology and Aging*, *31*(7), 758–770.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3), 1–46.

Bäuml, K.-h. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review*, *5*(3), 459–463.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *The Journal of General Psychology*, *30*(2), 149–165.

Brewer, G. A., Marsh, R. L., Clark-Foos, A., & Meeks, J. T. (2010). Noncriterial recollection influences metacognitive monitoring and control processes. *Quarterly Journal of Experimental Psychology*, *63*(10), 1936–1942.

Bruce, P. R., Coyne, A. C., & Botwinick, J. (1982). Adult age differences in metamemory. *Journal of Gerontology*, *37*(3), 354–357.

Brydges, C. R., & Bielak, A. A. (2020). A Bayesian analysis of evidence in support of the null hypothesis in gerontological psychology (or lack thereof). *The Journals of Gerontology: Series B*, *75*(1), 58–66.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*(5), 654–663.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633–642.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*(1), 50–71.

Cox, G. E., Hemmer, P., Aue, W. R., & Criss, A. H. (2018). Information and processes underlying semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General*, *147*(4), 545–590.

Coyne, A. C. (1985). Adult age, presentation time, and memory performance. *Experimental Aging Research*, *11*(3), 147–149.

Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science*, *3*(5), 155–158.

Craik, F. I. M., Morris, L. W., Morris, R. G., & Loewen, E. R. (1990). Relations between source amnesia and frontal lobe functioning in older adults. *Psychology and Aging*, *5*(1), 148–151.

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*(4), 316–326.

Daniels, K. A., Toth, J. P., & Hertzog, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology and Aging*, *24*(2), 494–500.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, *12*(1), 45–75.

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging*, *5*(2), 291–303.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, *11*(9), 379–386.

Diaz, M., & Benjamin, A. S. (2011). The effects of proactive interference (PI) and release from PI on judgments of learning. *Memory & Cognition*, *39*(2), 196–203.

Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology*, *38*(6), 682–688.

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, *22*(1), 122–133.

Dougherty, M. R., & Hunter, J. E. (2003a). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica*, *113*(3), 263–282.

Dougherty, M. R., & Hunter, J. E. (2003b). Probability judgment and subadditivity: The role of working memory capacity and constraining retrieval. *Memory & Cognition*, *31*(6), 968–982.

Duchek, J. M. (1984). Encoding and retrieval differences between young and old: The impact of attentional capacity usage. *Developmental Psychology*, *20*(6), 1173–1180.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (jols): Effects of relatedness and serial position on jols. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1180–1191.

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52*(4), 526–534.

Eakin, D. K. (2010). Listchecker Pro 1.2: A program designed to facilitate creating word lists using the University of South Florida word association norms. *Behavior Research Methods*, *42*(4), 1012–1021.

Eakin, D. K., & Hertzog, C. (2006). Release from implicit interference in memory and metamemory: Older adults know that they can't let go. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *61*(6), P340–P347.

Eakin, D. K., & Hertzog, C. (2012a). Age invariance in feeling of knowing during implicit interference effects. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *67*(5), 555–562.

Eakin, D. K., & Hertzog, C. (2012b). Immediate judgments of learning are insensitive to implicit interference effects at retrieval. *Memory & Cognition*, *40*(1), 8–18.

Eakin, D. K., Hertzog, C., & Harris, W. (2014). Age invariance in semantic and episodic metamemory: Both younger and older adults provide accurate feeling-of-knowing for names of faces. *Aging, Neuropsychology, and Cognition*, *21*(1), 27–51.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 1–9.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224.

Goodman, L. A., & Kruskal, W. H. (1963). Measures of association for cross classifications III: Approximate sampling theory. *Journal of the American Statistical Association*, *58*(302), 310–364.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 163–169.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of learning and motivation* (pp. 193–225). Elsevier.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance* (pp. 653–675). MIT Press.

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and aging. In T. Salthouse & F. I. M. Craik (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 417–466). Lawrence Erlbaum Associates.

Hertzog, C., & Curley, T. (2018). Metamemory and cognitive aging. Oxford University Press.

Hertzog, C., Curley, T., Castro, N., & Dunlosky, J. (2020). *Evidence of age differences in episodic feeling-of-knowing but not retrospective confidence judgments after distinctiveness-based encoding*, Unpublished manuscript, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia, USA.

Hertzog, C., Curley, T., & Dunlosky, J. (2021). Are age differences in recognition-based retrieval monitoring an epiphenomenon of age differences in memory? *Psychology and Aging*, *36*(2), 186–199.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*(2), 215–227.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 22–34.

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, *38*(6), 771–784.

Hertzog, C., Fulton, E. K., & Dunlosky, J. (2020). *Age differences in accuracy of feeling-of-knowing judgments are connected to noncriterial recollection of original encoding experiences*, Unpublished manuscript, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia, USA.

Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & Cognition*, *42*(1), 126–140.

Hertzog, C., Hülür, G., Gerstorf, D., & Pearman, A. M. (2018). Is subjective memory change in old age based on accurate monitoring of age-related memory change? evidence from two longitudinal studies. *Psychology and Aging*, *33*(2), 273–287.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging*, *17*(2), 209–225.

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, *46*(4), 939–948.

Hertzog, C., Small, B. J., McFall, G. P., & Dixon, R. A. (2019). Age, cohort, and period effects on metamemory beliefs. *Psychology and Aging*, *34*(8), 1077–1089.

Hess, T. M. (2014). Selective engagement of cognitive resources: Motivational influences on older adults' cognitive functioning. *Perspectives on Psychological Science*, *9*(4), 388–407.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(1), P3–P11.

Hicks, J. L., & Marsh, R. L. (2002). On predicting the future states of awareness for recognition of unrecallable items. *Memory & Cognition*, *30*(1), 60–66.

Higham, P. A., & Higham, D. P. (2019). New improved gamma: Enhancing the accuracy of Goodman–Kruskal's gamma using ROC curves. *Behavior Research Methods*, *51*(1), 108–125.

Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. *The Oxford Handbook of Metamemory*, 39–61.

Hines, J. C., Hertzog, C., & Touron, D. R. (2015). Younger and older adults weigh multiple cues in a similar manner to generate judgments of learning. *Aging, Neuropsychology, and Cognition*, *22*(6), 693–711.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hulbert, J. C., Shivde, G., & Anderson, M. C. (2012). Evidence against associative blocking as a cause of cue-independent retrieval-induced forgetting. *Experimental Psychology*, *59*(1), 11–21.

Ikier, S., & Hasher, L. (2006). Age differences in implicit interference. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *61*(5), P278–P284.

Ikier, S., Yang, L., & Hasher, L. (2008). Implicit proactive interference, age, and automatic versus controlled retrieval strategies. *Psychological Science*, *19*(5), 456–461.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, *17*(1), 3–11.

Kausler, D. H. (1994). *Learning and memory in normal aging*. Academic Press San Diego, CA.

Kay, H., & Poulton, E. C. (1951). Anticipation in memorizing. *British Journal of Psychology*, *42*(1), 34–41.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and language*, *35*(2), 157–175.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language*, *48*(4), 704–721.

Koen, J. D., & Yonelinas, A. P. (2016). Recollection, not familiarity, decreases in healthy ageing: Converging evidence from four estimation methods. *Memory*, *24*(1), 75–88.

Koop, G. J., Criss, A. H., & Malmberg, K. J. (2015). The role of mnemonic processes in pure-target and pure-foil recognition memory. *Psychonomic Bulletin & Review*, *22*(2), 509–516.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 34–53.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478–492.

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica*, *58*(2), 141–158.

Lachman, J. L., Lachman, R., & Thronesbery, C. (1979). Metamemory through the adult life span. *Developmental Psychology*, *15*(5), 543–551.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *Sage Open*, *6*(1), 1–17.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, *42*(1), 333–376.

Lineweaver, T. T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition*, *5*(4), 264–296.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199–207.

MacLaverty, S. N., & Hertzog, C. (2009). Do age-related differences in episodic feeling of knowing accuracy depend on the timing of the judgement? *Memory*, *17*(8), 860–873.

Maki, R. H. (1999). The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1011–1023.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430.

Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509–527.

Mazancieux, A., Dinze, C., Souchay, C., & Moulin, C. J. (2020). Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence. *Neuroscience of Consciousness*, *6*(1), 1–11.

McEvoy, C. L., Holley, P. E., & Nelson, D. L. (1995). Age effects in cued recall: Sources from implicit and explicit memory. *Psychology and Aging*, *10*(3), 314–324.

Metcalfe, J. (2000). Feelings and judgments of knowing: Is there a special noetic state? *Consciousness & Cognition*, *9*(2), 178–186.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851–861.

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for common designs* [R package version 0.9.12-4.2].

Morson, S. M., Moulin, C. J., & Souchay, C. (2015). Selective deficits in episodic feeling of knowing in ageing: A novel use of the general knowledge task. *Acta Psychologica*, *157*, 85–92.

Moscovitch, M., & Winocur, G. (1992). The neuropsychology of memory and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 315–372). Lawrence Erlbaum Associates.

Narens, L., Nelson, T. O., & Scheck, P. (2008). Memory monitoring and the delayed JOL effect. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 137–153). Taylor Francis.

Neely, J. H., Schmidt, S. R., & Roediger, H. L. (1983). Inhibition from related primes in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(2), 196–211.

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition*, *41*(6), 797–819.

Nelson, D. L., & McEvoy, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 292–314.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1990). Encoding context and retrieval conditions as determinants of the effects of natural category size. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 31–41.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*(2), 299–324.

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review*, *99*(2), 322–348.

Nelson, D. L., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. *Psychonomic Bulletin & Review*, *7*(4), 604–617.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science*, *2*(4), 267–271.

Nelson, T. O., Leonesio, J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 279–288.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of learning and motivation* (pp. 125–173). Elsevier.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Perfect, T. J., & Dasgupta, Z. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition*, *25*(6), 849–858.

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults: One deficit or two? *The Quarterly Journal of Experimental Psychology*, *46*(1), 119–135.

Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology*, *14*(4), 330–345.

Perrotin, A., Isingrini, M., Souchay, C., Clarys, D., & Taconnat, L. (2006). Episodic feeling-of-knowing accuracy and cued recall in the elderly: Evidence for double dissociation involving executive functioning and processing speed. *Acta Psychologica*, *122*(1), 58–73.

Perrotin, A., Tournelle, L., & Isingrini, M. (2008). Executive functioning and memory as potential mediators of the episodic feeling-of-knowing accuracy. *Brain and Cognition*, *67*(1), 76–87.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, *7*(3), 672–694.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435–451.

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. *The Oxford Handbook of Metamemory* (pp. 1–19). Oxford University Press.

Robinson, A. E., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. *Aging, Neuropsychology, and Cognition*, *13*(3-4), 458–478.

Roediger, H. L. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 644–657.

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*(2), 261–269.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*(9), 1043–1056.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 91–105.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, *22*(5), 511–524.

Rosen, A. C., Prull, M. W., O'Hara, R., Race, E. A., Desmond, J. E., Glover, G. H., Yesavage, J. A., & Gabrieli, J. D. (2002). Variable effects of aging on frontal lobe contributions to memory. *NeuroReport*, *13*(18), 2425–2428.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 43–50.

Sacher, M., Isingrini, M., & Taconnat, L. (2013). Effects of aging and divided attention on episodic feeling-of-knowing accuracy. *Acta Psychologica*, *144*(2), 258–263.

Sacher, M., Landré, L., & Taconnat, L. (2015). Age-related differences in episodic feeling-of-knowing arise from differences in memory performance. *Memory*, *23*(2), 119–126.

Salthouse, T. A. (1992). Why do adult age differences increase with task complexity? *Developmental Psychology*, *28*(5), 905–918.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 39–54.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 501–518.

Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle Longitudinal Study*. Oxford University Press.

Schreiber, T. A. (1998). Effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory & Cognition*, *26*(3), 553–571.

Schreiber, T. A., & Nelson, D. L. (1998). The relation between feelings of knowing and the number of neighboring concepts linked to the test cue. *Memory & Cognition*, *26*(5), 869–883.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 452–460.

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 400–408.

Smith, A. D. (1975). Aging and interference with memory. *Journal of Gerontology*, *30*(3), 319–325.

Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Souchay, C., & Isingrini, M. (2004a). Age related differences in metacognitive control: Role of executive functioning. *Brain and Cognition*, *56*(1), 89–99.

Souchay, C., & Isingrini, M. (2004b). Age-related differences in the relation between monitoring and control of learning. *Experimental Aging Research*, *30*(2), 179–193.

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, *14*(2), 299–309.

Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, *16*(4), 769–784.

Taub, H. A., & Walker, J. B. (1970). Short-term memory as a function of age and response interference. *Journal of Gerontology*, *25*(3), 177–183.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2011). Context affects feeling-of-knowing accuracy in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 96–108.

Thomas, R. P., Dougherty, M. R., & Buttaccio, D. R. (2014). Memory constraints on hypothesis generation and decision making. *Current Directions in Psychological Science*, *23*(4), 264–270.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*(1), 155–185.

Toth, J. P., Daniels, K. A., & Solinger, L. A. (2011). What you know can hurt you: Effects of age and prior knowledge on the accuracy of judgments of learning. *Psychology and Aging*, *26*(4), 919–931.

Toth, J. P., & Parks, C. M. (2006). Effects of age on estimated familiarity in the process dissociation procedure: The role of noncriterial recollection. *Memory & Cognition*, *34*(3), 527–537.

Tulving, E. (1989). Memory: Performance, knowledge, and experience. *European Journal of Cognitive Psychology*, *1*(1), 3–26.

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*(5), 321–334.

Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*(1), 145–150.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, *50*(3), 289–335.

Wahlheim, C. N. (2011). Predicting memory performance under conditions of proactive interference: Immediate and delayed judgments of learning. *Memory & Cognition*, *39*(5), 827–838.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442–452.

Wickelgren, W. A. (1976). Memory storage dynamics. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (pp. 321–361). Erlbaum.

Wickens, D. D. (1973). Some characteristics of word encoding. *Memory & Cognition*, *1*(4), 485–490.

Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *The American Journal of Psychology*, *114*(3), 329–354.

Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, *95*, 78–88.

Wilson, J. H., Kellen, D., & Criss, A. H. (2020). Mechanisms of output interference in cued recall. *Memory & Cognition*, *48*(1), 51–68.

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1024–1039.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*(1), 89–106.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition*, *5*(1-2), 131–141.

Zogg, J. B., Woods, S. P., Sauceda, J. A., Wiebe, J. S., & Simoni, J. M. (2012). The role of prospective memory in medication adherence: A review of an emerging literature. *Journal of Behavioral Medicine*, *35*(1), 47–62.