EFFECTS OF CONTEXT AND AGE ON FEELING OF KNOWING (FOK) JUDGMENTS

by

TASNUVA ENAM

IAN M. MCDONOUGH, COMMITTEE CHAIR

REBECCA S. ALLEN
SHEILA BLACK
JOANNA BARTH
ED MERRILL

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2020

ABSTRACT

In two experiments we wanted to examine how does age and context influence FOK ratings and recognition memory. In Experiment 1 we found age and context have no influence on FOK and recognition memory but influences subjective memory ratings such as recollection and familiarity to a greater extent. Older adults specifically seem to have higher subjective ratings for recollection and familiarity between old and new scenes compared to younger adults. In Experiment 2, explicit instruction was added to draw attention away from context. We found with explicit instruction, older adults have higher ratings overall for FOK judgments and higher subjective memory ratings for recollection and familiarity compared to younger adults. Results may help us to understand how FOK judgments are made as we age in addition to importance of contextual information in subjective memory in aging.

DEDICATION

This work is dedicated to my Amma and Abba—Thanks for being my pillar throughout all my good and bad times and reminding me that I have wings to fly :)

LIST OF ABBREVIATIONS AND SYMBOLS

a	Cronbach's index of internal consistency
β	Beta: standard regression coefficient
F	Fisher's *F* ratio: A ration of two variances
M	Mean: the sum of a set of measurements divided by the number of measurements in the set
N	Sample Size
η^2	Eta squared: measure of effect size
p	Probability associated with the occurrence under the null hypothesis of a value as extreme as or more extreme than the observed value
r	Pearson product-moment correlation
SD	Standard deviation: value of variation from the mean
<	Less than
>	Greater than
=	Equal to
Ɛ	Epsilon
χ^2	Chi-Square

ACKNOWLEDGMENTS

I would like to thank all the people who supported me to get this project completed. This includes my graduate advisor, Dr. Ian M. McDonough, my committee members (last names alphabetized), Dr. Rebecca S. Allen, Dr. Sheila Black, Dr. Joan Barth, Dr. Ed Merrill, and my research assistants, Rebekah, Gabrielle, and David. Thanks to the UA Subject Pool, ARIA database, Tuscaloosa, and Birmingham Community in general for helping me with the recruitment process. Last but not least, special thanks go to my awesome friends, Jessica and Janet, for giving me the mental support I needed to get this far~

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

A recent 2020 survey from the US Census Bureau (Medina, Sabo & Vespa, 2020) informs us that we as human beings of all races, gender, and ethnic groups are expected to live longer. While this is good news, living longer also comes with a price. In addition to a decline in physical health, aging also impacts cognitive health. One of the most debilitating cognitive declines that also influence physical health in aging is a decline in memory. Poor memory can not only be debilitating to self but also project to the community in form of ageism, a negative stereotype people have on older adults, which can then lead to discrimination (Butler, 2010). For our younger generation, caring about age-related issues is critical to address because eventually we will also become old and may face the same prejudice and discrimination current older adults face in form of ageism. Thus, as a community, we must work towards building an age-friendly community. One of the initial steps towards our goal would be to understand and address problems older adults face as they age. One of these issues is the memory deterioration in old age. While there is much research showing memories worsen with aging, fortunately, there are also studies that show older adults can use many kinds of cues to buffer their memory errors or even improve it. One such cue that acts as a memory aid for older adults is context or background scenes (Memel & Ryan, 2017; Burke et al., 2018).

Context is any kind of background information (Reder et al., 2013; Smith & Manzano, 2010; Smith & Vela, 2001) that is peripheral to the focus of our attention. Context of any stored memory may include details of the surrounding environment where the learning occurred.

Context influences how we make our memory for focal items in the context. In general, we never encode an object by itself, rather we encode some background stimuli to surrounding the object in hope that perhaps later, it will remind us of the focal object (Bar, Aminoff, & Schacter, 2008). Additionally, context plays a critical role in how older adults make their memory. As will be discussed in detail in the later sections, older adults rely more on context compared to younger adults due to their age-related deficiency in encoding memory details (Memel & Ryan, 2017; Burke et al., 2018). Thus, context for aging acts as a supportive memory aid in older adults. Older adults are also poor at perceiving their memory deficits that influences what strategies they can take to improve their actual memory. For example, if an older adult is not aware that their memory is bad, how can they know when to take steps to improve their actual memory? Most previous research has focused on all these factors-context, aging, and memory perceptions, but individually. In the current study, we examine if we can improve these miscalibrated memory perceptions and subsequent memory in older adults by using contextual cues they already find helpful.

In the current study, we operationalize our environmental background context similar to the previous memory studies which also defined their context as details of the surrounding environment when the learning took place, such as background photographs or scenes of places, background music, or short films as live scenes (Murrane, Phelps, & Malmberg, 1999; Reder et al., 2013; Smith & Manzano, 2010; Smith & Vela, 2001). Although there are some variations in previous studies on what kind of contexts are presented, such as related or unrelated context to the focal item, in the current study context definition is extended to the idea that we are not meant to pay attention to context. Thus, instead of providing related context to the focal item, our context will be unrelated to the focal item. The literature review below is divided into sections

with first starting with the beneficial effects of context, memory errors that occur as we age, how context influences memory and subsequent errors, and finally, transitioning to the context-metamemory relationship.

Influence of Context in Learning and Memory

Contexts, in general, have many practical applications in everyday life in terms of processing some new information and retrieval of that learned information (Cutler et al., 1987; Krafka et al., 1985; Craik & Schloerscheidt, 2011; Fernandez & Alonso, 2001; Naveh-Benjamin, & Craik, 1995; Smith & Vela, 1992; Smith, Vela, & Williamson, 1988). One example could be in eyewitness testimony. Returning the witness to that same context of the crime (i.e. reinstating the context) may help them witnesses to retrieve accurate details of the crime and identify the perpetrator (Tulving, 1983; Cutler et al., 1987; Krafka et al., 1985; Smith & Vela, 1992). In this example, the crime or the perpetrator can be the focal memory items and the background environment that the crime happened could be the context. Another example could be in learning. Quality of retrieving learned materials might differ depending on the change in the original background environment where learners learned the information (Godden & Baddeley, 1975; Smith & Vela, 2001). The retrieval of the learned material could be the focal memory item and the environment manipulation could be the context. This kind of context learning may also be important in academic settings, such as for students and for educators since both parties would want independent learning to take place without any dependency on classroom context or background context. Similarly, context may help age-related memory deficits in older adults by providing environmental background information which acts as environmental support to aid in their memory recall (Craik, 1983; Craik & Schloerscheidt, 2011; Fernandez & Alonso, 2001;

Naveh-Benjamin & Craik, 1995). Thus, the importance of integrating context is not only limited for the aging groups, but across ages and all scenarios.

Memory Measures and Aging

Memory retrieval is not always perfect and one of the ways to understand memory failure is to investigate the memory errors people make in recognition memory. In recognition memory tests, participants have to discriminate between studied and non-studied items (Yonelinas, 1999). False memory occurs when participants incorrectly believe they have seen a new item before. False memory in recognition is usually measured via old/new recognition judgment task. Participants have to identify whether they have seen the particular stimuli in the past and respond "old" or if they have never seen the stimuli before and respond "new". The task results in 4 types of responses: hits, misses, false alarms, and correct rejections. From these responses, we get two types of calculations that are used for final analysis: hit rates, defined as correctly recognizing the target, and are calculated as (1-number of misses) and false alarms, defined as incorrect recognition of lure and calculated as (1-correct rejection). While hit rates only show us what was recalled, it provides no information on the ability of the subject to discriminate between old and new items. Thus, we must also look at the false alarm rate, the proportion of times that the subject said New, but the test item was not on the list, essentially their false memory (Table 1).

Table 1

Old/New Judgment Responses

	Respond "Old"	Respond "New"
Target	Hits	Misses
Lure	False Alarm	Correct Rejection

Recollection-Familiarity: In addition to our objective memory, subjective memory also plays a role to improve objective memory since the deliberate conscious effort to improve our objective memory is likely to occur if we have the perception to have memory deficits in the first place (Ba¨ckman & Dixon, 1992). Subjective memory plays a key role in the aging process by compensating or buffering age-related deficits (Garett, Grady & Hasher, 2010). Two key factors, recollection, and familiarity influence false alarm rates for recognition memory. According to the dual-process model of memory, we evaluate our memory events with two types of subjective memory processes: recollection, which refers to a more effortful, conscious process that involves the recall of actual details of a past item, and familiarity, which refers to an automatic process that provides immediate information about whether the current item has been encountered before (Yonelinas, 2002). Previous research has shown that generally older adults have intact familiarity as younger adults but have poor recollection ability compared to younger adults (Jennings & Jacoby, 1993; Yonelinas, 2002).

Remember-Know: Recollection and familiarity are usually studied by “remember” (R) versus “know” (K) procedure, where at cued recall participants are instructed to respond to “R” if they can recall the item with accurate details or “K” if they can recall the item without any details (Java, 1996; Yonelinas, 2002). The procedure allows evaluating whether, at cued recall, participants are recalling the items with recollection-based processes that are defined with the ‘remember” responses or familiarity-based process, which are defined with the “know” responses. Previous studies have shown that older adults have poor R responses compared to younger adults while no age impairment was found in K responses (Java, 1996; Mäntylä, 1993; Yonelinas, 2002). However, for false memories, recollection and familiarity are defined more specifically.

Distinctive Heuristic and False Memory: Previous research has shown that false memory can also be reduced with distinctive heuristic processes where participants reject a source of being previously studied by failing to recollect expected information related with the source (McDonough & Gallo, 2008; Schacter, Israel, & Racine, 1999). Previous studies have found that false recollection in aging can be reduced further with picture stimuli because pictures are thought to provide more distinctive heuristic (Craik & Schloerscheidt, 2011; Eakin & Hertzog, 2012; Eakin, Hertzog, & Harris, 2015; Gutchess & Park, 2009; Luo, Hendriks, & Craik, 2007; MacLaverty & Hertzog, 2009; Nelson, Reed, & McEvoy, 1977; Park, Puglisi, & Sovacool, 1983). The pictorial advantage in retrieval process is known as the picture superiority effect and it occurs since pictures provide greater elaborative encoding, which is thought to improve recollection and in turn reduce the burden of associative binding, especially in older adults (Craik & Schloerscheidt, 2011; Eakin & Hertzog, 2012; Eakin, Hertzog, & Harris, 2015; Gutchess & Park, 2009; Luo, Hendriks, & Craik, 2007; Park, Puglisi, & Sovacool, 1983; MacLaverty & Hertzog, 2009). The distinctive heuristic model is also connected within the dual-process model by defining recollection and familiarity more specifically at old/new recognition memory tests (Gallo, Weiss, & Schacter, 2004).

Recollection is defined as the retrieval of information by recalling the target from the retrieval cue. Particularly, when participants see a retrieval cue, such as a non-studied picture, the retrieval cue has to bring to mind the target picture and all the unique information about the target when it was paired with the retrieval cue (Gallo, Weiss, & Schacter, 2004; McDonough & Gallo, 2008; Schacter, Israel, & Racine, 1999). If the retrieval cue does not bring to mind any recollection-based information about the target, then participants can reject the retrieval cue as not being associated with the target before and therefore define the item as a “new” item.

Familiarity, on the other hand, is defined as a vague feeling that the retrieval cue was processed without bringing into any mind any specific details on the target item (Gallo, Weiss, & Schacter, 2004; McDonough & Gallo, 2008; Schacter, Israel, & Racine, 1999). Making judgments based on distinctive heuristic are more diagnostic because participants correctly reject a lure based on information they fail to recollect about the target, based on that lure. Instead of R vs. K procedure at cued recall, participants use a "Recollection" vs. "Familiar" procedure with similar instructions, but at recognition tests.

Previous research has shown older adults have reduced recollection, but intact familiarity responses compared to younger adults, which might indicate that distinctive heuristic mechanism of older adults with pictures and contexts might be comparable to younger adults (Gallo, Weiss, & Schacter, 2004; McDonough & Gallo, 2008; Schacter, Israel, & Racine, 1999; Yonelinas, 2002). For example, similar lures may bring about a sense of "familiarity" however, participants, including older adults, might still correctly reject the lure based on the absence of that distinct property from the studied picture.

On the other hand, the inability to ignore irrelevant context at study may reduce older adults' memory performance compared to younger adults. This may make them rely more on their intact familiarity and result in higher false alarms to lures (Gallo, Weiss, & Schacter, 2004; McDonough & Gallo, 2008; Schacter, Israel, & Racine, 1999; Yonelinas, 2002). The distinctive heuristic model is mostly used for old/new recognition memory tests and is used to distinguish between false alarms for studied items and similar lures. In the current study, we will employ a similar mechanism to investigate false memory in aging that may arise due to different context conditions.

Recall-To-Reject Strategy and False Memory: Beside distinctive heuristics, participants can also simply use a recall-to-reject strategy where they recall the correct pair of items *first* (unlike in distinctive heuristic where they fail to recollect the information) and then reject the source if the recalled information does not match with the retrieval source (Gallo, Bell, Beier, & Schacter, 2006). Thus, recall to reject strategy is based on logically inconsistent information (e.g., ‘‘I didn’t see this pair, orange-pencil because I remember seeing apple-cat, and none of the pairs in the study list had any picture of an orange or a pencil ’’) unlike distinctive heuristic, which is based on retrieval source information (e.g., ‘‘I didn’t see this pair of orange-pencil because I would remember it if I had’’) (Gallo, Bell, Beier, & Schacter, 2006). Both processes help to reduce false memories.

Norman and Schacter (1997) suggested that in general, older adults are less likely than younger adults to strategically evaluate their memory, such as using recall to reject strategies and thus are more likely to respond “New” to lures, resulting in higher false alarms than younger adults. Moreover, in general, older adults also have impaired associative memory, filling in missing information from episodic memory (pattern completion), failure in separating similar pieces of information (pattern separation) and poor processing speed (Allen-Burge and Storandt, 2000; Chalfonte et al., 1997; Bakker et al., 2008). If older adults are unable to suppress irrelevant context at study, then it will also result in false alarms in recognition. Overall, many different factors, poor recollection, poor binding ability (discussed more in later sections), impaired associative memory, may interact to influence retrieval processes at recognition memory. In the next section, we talk about how context and age can influence recognition memory.

Recognition Memory

Older adults usually have equivalent memory performance to younger adults, *but only* if the context is *related* to the focal items (Murnane, Phelps, & Malmberg, 1999; Hanczakowki, Zawadska, & Coote, 2014; Smith & Vela, 2001). This is not surprising as previous studies have shown related context provide semantic meaning to the focal items and result in higher memory accuracy. Examples include instances when context at test was reinstated from the encoding phase, when picture item and picture context was related, or when target words were presented in a sentence context. In all these examples, older adults recognition memory for items were equivalent to younger adults (Cohen & Faulkner, 1983; Craik & Schloerscheidt, 2011; Eakin & Hertzog, 2012; Eakin, Hertzog, & Harris, 2015; Gutchess & Park, 2009; Luo, Hendriks, & Craik, 2007; MacLaverty & Hertzog, 2009; Park, Smith, Morell, Puglisi, & Dudley, 1990; Park, Puglisi, Smith, & Dudley, 1987; Park, Puglisi, & Sovacool, 1984).

Previous work on context has shown reinstated context increases recognition accuracy by increasing hits to studied items and reducing false alarms to lures (Murnane, Phelps, & Malmberg, 1999; Hanczakowki, Zawadska, & Coote, 2014). Craik and Jennings (1992) and Park et al. (1990) in their context integration hypothesis for aging proposed that when contexts are related to the focal items, older adults might sometimes be even more facilitated than younger adults, thus overall showing sometimes context can act as a supportive aid in older adults to improve their recognition memory and hinting that it may help reduce false memory. All these studies also support the encoding specificity and the context reinstatement hypotheses. Reinstating context for memory improvement is guided by the encoding specificity theory, which states that memory accuracy is higher when there is an overlap between cues at the encoding and retrieval stage (Smith & Vela, 2001; Tulving & Thompson, 1973). Thus, according

to context reinstatement theory, it also implies reinstating original context result in accurate recognition memories.

Previous research has also studied the effects of distracting context indirectly through divided attention tasks and have found that doing a secondary task with the main task reduces older adults' working memory (Baddeley & Hitch, 1974; McDowd & Craik, 1998). In divided attention, participants usually study a list of stimuli for later memory tasks and at the same time have to do another task in the background. From a stimuli-context viewpoint, the main task could be the focal item while the second task could be the context. Some earlier studies have shown that divided attention does not produce an age difference between younger and older adults if the focal task itself is very simple (Park, Puglisi, Smith, & Dudley, 1987). Other studies have found age differences with increasing task complexity, showing that older adults perform poorly than younger adults, such as making higher false alarms for old items or re-paired conjunction of word pairs, under divided attention in recognition judgments (Castel & Craik, 2003; Jacoby & Jennings, 1993; McDowd & Craik, 1998; Park, Smith, Dudley, & Lafronza, 1989). One reason for older adults' reduced performance could be it involves the allocation of limited resources (Salthouse et al., 1984). Thus, overall, there seems to be contrasting evidence inferred from divided attention literature on how distracting context affects different age groups.

It should be noted that in divided attention, participants are asked to pay attention to the secondary task because they also must keep track of it. In this sense, the task competes with the primary task and is not the best way to define a context. Unrelated contexts, as in the current study, are background context that is peripheral to the focus of the attention. One way to test for effects of distracting contexts between age groups would be to investigate the ability to ignore

irrelevant contexts such as unrelated contexts in the current study (Experiment 2, described in a later section in Methodology) such as in distractor task designs. Previous research has proposed that older adults may have reduced inhibiting the ability to ignore irrelevant information (Hasher & Zachs, 1988). The next section details how inhibition or suppression ability for irrelevant information changes with aging.

Inhibition Mechanism

Older adults have reduced ability in suppressing irrelevant information and besides, tend to store irrelevant information in their working memory (Hasher & Zachs, 1988), which in turn reduces their ability to process focal targets. Specifically, age-related inhibition in focal memory tasks is only apparent when a dominant response is required for secondary tasks (Rey-Mermet & Gade, 2017). However, these studies only test working memory in general but do not distinguish between phases of memory, specifically at retrieval which is more prone to memory decline with age (Yonelinas, 2002). In studies with irrelevant contexts that are not associated with target items, research from cued recall has shown that generally, older adults' encoding ability for target items decline with irrelevant context which in turn decline their overall memory performance compared to younger adults (Earls, Smith & Park, 1994).

A recent review has also shown that instead of only reduced inhibitory ability shown in previous studies, age effects shown in inhibitory tasks could be due to different types of experimental tasks or the level of executive function in older adults (Rey-Mermet & Gade, 2017). Contrasting evidence has been found where older adults with higher executive function seldom showed age discrepancy in inhibitory tasks compared to younger adults and if there was an age discrepancy, then the tasks were those that required coordinating and updating

information such as Simon or Stroop Task. This suggests perhaps, in addition to the only decline in inhibition, older adults also have reduced ability in performing complex tasks with irrelevant information. If such is the case, then complex tasks that require distinguishing between old and new items, monitoring memory, the decision-making process for future memory may also show an age-related decline. In addition to inhibition, aging also reduces our binding mechanisms for associative memory, as described in the next sections.

Binding Mechanism

Binding mechanism is a complex function required to associate between different elements from past correctly to form a coherent memory episode and this ability is needed in retrieval for recognition memory (Chalfonte & Johnson, 1996; Naveh Benjamin, 2000; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003; Naveh-Benjamin, Guez, Kilb, & Reedy, 2004; Park et al., 1990, Yonelinas, 2002). Impaired binding ability result in lower accuracy in discriminating between old and new items in recognition memory tasks. Older adults generally have also shown poor binding ability resulting in poor associative memory (Chalfonte & Johnson, 1996; Naveh Benjamin, 2000; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003; Naveh-Benjamin, Guez, Kilb, & Reedy, 2004; Park et al., 1990, Yonelinas, 2002). Poor binding ability are evident in studies where participants have to associate target-context and also have to discriminate between conceptually *similar lures*, lures that are similar to the study items but have never been shown before in the study phase, and *reinstated contexts*, contexts they have seen before in the study phase (Doss, Maximilan, Bluestone, & Gallo, 2016; Doss, Picart, & Gallo, 2018). For example, when participants saw similar lures to items on a reinstated context at test, they were more likely to confuse conceptual aspects of similar lures and studied items, and in turn created false

memories of context-lure association (Doss, Maximilan, Bluestone, & Gallo, 2016; Zaragoza, Mitchell, Payment, & Drivdahl, 2011).

The results imply to the idea that older adults may have reduced ability in discriminating between old and new items, between similar items and binding specific aspects to form a coherent episode to retrieve accurately at retrieval. Additionally, older adults, in general, has also been shown to be impaired in tasks that require filling in missing information from episodic memory (pattern completion) or failure in separating similar pieces of information (pattern separation) (Chalfonte et al., 1997; Bakker et al., 2008). Doss et al. (2018) explained the effect of context reinstatement via their context distortion hypothesis where it is the aspect of lures, particularly similar lures that are conceptually similar to the studied items, combined with the aspect of relating target-context, that increases false memories at context reinstatement because participants confuse the conceptual binding between similar lures-target items and context. Research using neuroimaging has shown that on average older adults make higher false alarms than younger adults when they see old items on a new background or a new object on an old background. Referring to the Doss et al. (2018) hypothesis, even with irrelevant context, older adults may be more prone to be confused about the conceptual similarity between similar lures-target pairs and context in a reinstated context condition.

So far, we talked about objective memory that impairs with age. Although research from subjective memory tells us that implicit memory (Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004), in this case, familiarity, is reserved with age, research from both behavioral and neuroimaging evidence shows that this implicit memory also increases for irrelevant and distracting items (Hasher, Zacks, & May, 1999; Rowe et al., 2006). In an fMRI task (Campbell,

Grady, Ng, & Hasher, 2012), older and younger adults were given a 1-back task with some target words superimposed with non-words with an object. The next day they were tested for their implicit memory across words with some actual target words and some non-words without them being aware. Older adults had higher implicit memory for the non-words compared to younger adults. The greater distracted ability due to age is thought to occur due to reduced connectivity in the frontoparietal attentional network that normally reduces attention to distracting information during the encoding phase of our memory (Campbell, Grady, Ng, & Hasher, 2012; Rowe, Valderrama, Hasher, & Lenartowicz, 2006; Weeks & Hasher, 2014; Hasher, Zacks, & May, 1999; Rowe et al., 2006). Thus, just as there is an age-related deficiency in binding correct target information, the phenomenon of binding irrelevant information to target due to distracted attention is also prone to aging and is known as hyper-binding. (Campbell, Hasher, & Thomas, 2010). As mentioned above most studies have shown negative effects of hyper-binding in memory (Campbell, Grady, Ng, & Hasher, 2012; Rowe, Valderrama, Hasher, & Lenartowicz, 2006; Weeks & Hasher, 2014; Hasher, Zacks, & May,1999; Rowe et al., 2006).

But hyper-binding to distractors in an implicit manner can also have beneficial effects. Participants subsequently report not paying any conscious attention to distractors. For example, previous research has shown that hyper-binding can help older adults learn target pictures and simultaneous distractor names in the face-name associative task (Weeks, Biss, Murphy & Hasher, 2016). Hyper-binding can also help older adults when distractor items at encoding become a to-be-remembered item in later memory task (Thomas & Hasher, 2012) or when distraction serves as a rehearsal opportunity between encoding and later test (Biss, Ngo, Hasher, Campbell, & Rowe, 2013). Thus, context can act both as a supportive aid and a distractor during a memory task in aging.

Feeling of Knowing (FOK) Judgments

Although extensive research has examined how aging and context can influence recognition memory, less has been focused on another type of memory process that controls our encoding and retrieval strategies in making such memories. This process is called metamemory. Metamemory is the process of judging and evaluating our memory and related strategies (Nelson & Narrens, 1990). Investigating metamemory judgments are a necessity since for one to have a good memory, one should also have an efficient metamemory system. We want to focus on a retrieval based metamemory judgment known as Feeling of Knowing (FOK) judgments, which are heavily dependent on our recognition memory systems.

FOK judgments are crucial in situations where people need to evaluate current knowledge, their confidence in the accuracy of retrieved information, or the prediction of future memory for an item-*a feeling of knowing*- that it can be retrieved later (Koriat, 1993; Nelson & Narrens, 1999). FOK judgments are different from TOT (tip-of the tongue-phenomenon) in that TOT access the current knowledge of unrecalled items participants might have from cued recall tests whereas FOK judgments access the current knowledge for future recognition memory performance (Hart, 1965). In this sense, FOK is more accurate compared to TOT since unlike cued recall, recognition memory is more sensitive in distinguishing between items. Unlike the TOT phenomenon, in doing FOK judgments, people still have an accurate prediction for the likelihood of correct recognition of an item that was just tested by the cued recall (e.g., Hart, 1967). Even when cued recall fails, FOK judgments demonstrate that evidence about whether the sought-after target is available in memory can still be accessed through recognition memory (e.g., Koriat, 2000; Nelson, Gerler, & Narens, 1984). In the paradigm developed by Hart (1965),

FOK judgments ask participants to predict their subsequent memory that participants cannot recall. In other words, TOT is more of a phenomenon that we know the information currently from cued recall but cannot produce it at the moment whereas, in FOK judgments, we are sure we will retrieve the information in subsequent recognition memory. Additionally, earlier studies only included unrecalled items for FOK judgment analysis, but in the present research methodologies, analyzing both recalled and unrecalled items is important as we cannot be sure that things we remember currently can be recalled in the future as well. Overall, in the current study, our FOK is defined as judgments where participants will predict future memory performance and for both recalled and unrecalled items.

Unlike other metamemory judgments that are immediately attempted before or after seeing the item, the paradigm for FOK judgments gives participants time between seeing an item and judging whether they will retrieve it later. This paradigm allows participants' FOK judgment processes akin to the real-life self-evaluation of their memory of an item. FOK judgments are related to various decision-making processes such as identifying which items are close to being learned and need to be restudied (Singer & Tiede, 2008; Hanczakowski, Zawadzka, & Cockcroft-McKay, 2014). FOK judgments consist of two components—feeling of familiarity of the cue-target and partial information of the target, both of which influences FOK judgments and retrieval of future memory in both younger and older adults (Koriat, 1993; Metcalfe, Schwartz, & Joaquim, 1993; Reder, 1987; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007; Schwartz & Metcalfe, 1992; Yonelinas, 2002). Since objective and subjective memory is influenced by age and context as mentioned in the early sections and since FOK judgments are dependent on both, likely, FOK judgments might also be prone to be influenced by age and context manipulation.

FOK paradigm and measures: In laboratory settings, FOK judgments are often studied by asking participants to study a list of word pairs, followed by a cued recall test, a FOK judgment phase, and a recognition test. For example, a cue-target word pair might be presented such as "cat-apple," where "cat" is the cue, "apple" is the target, followed by a cued recall test where the cue "cat" is presented, and participants have to recall the target. A cued recall phase is followed by a FOK judgment phase where participants are asked how likely they think they will remember the pairs when given the cue on a scale from 0 (will not remember) to 100 (will remember). Finally, participants are given a recognition test in which cues are presented with a set of lures, and participants must choose the correct target.

FOK Magnitude and FOK Accuracy: FOK magnitude is the average of mean FOK judgment ratings. Accurate metamemory magnitude would occur if higher metamemory judgments were associated with correct memory responses on a later memory test and if lower metamemory responses were associated with incorrect memory responses on a later memory test. FOK accuracy, also known as FOK resolution would be calculated using Goodman and Kruskal's gamma (G) correlation (Nelson, 1984). G correlation is a measure of the strength and direction of association that exists between FOK judgments and recognition that accounts for item and participant inter individuality and is a common measure used in metamemory literature. High FOK gamma score indicates good FOK ability and low FOK gamma score indicate poor FOK ability (Nelson, 1984). It is possible to have high FOK accuracy (good ability in predicting memory recognition) at various levels of recognition memory performance (Nelson, 1984). In the current study, we are interested in understanding *how* participants make the predictions itself and what influences their prediction ratings, not whether it correlates with their actual memory performance (accuracy/resolution/gamma scores). In other words, we are interested in examining

mean FOK ratings/calibration instead of gamma scores. Overall, this section is a summary of the feeling of knowing judgments and how researchers measure such predictions.

Although most studies have looked at how context influences recognition memory, very few have investigated how contexts influence FOK judgment ability independent of recognition memory and FOK judgment magnitude, both of which may influence the rate of false memories in subsequent memory tests. Reinstating context has been shown to enhance familiarity and retrieval of target information, both of which are components of FOK judgments that enhances not only the judgment ratings but also retrieval accuracy, mostly with recollection (Koriat, 1993; Macken, 2002). In younger adults, reinstating the original context has been shown to elevate the feeling of FOK judgments and confidence judgments, in other words, the mean rating, while re-paired context decreased FOK mean rating, but interestingly both had no effect on judgment accuracy, in other words, *the ability* to predict future recognition performance (Hanczakowski, Zawadzka, & Coote, 2015; Hanczakawoski, Zwadska, & Macken, 2017) showing that contexts do elevate the confidence of FOK judgment ratings but perhaps does not influence the ability to predict for future memory, at least in younger adults.

Research on the effect of context on FOK judgments in older adults is scarce. Older adults, on average, have associative binding deficits, poor recollection, the deficit in distinguishing similar patterns and combining reinstated memory accurately at retrieval (Bakker et al., 2008; Chanfolte & Johnson, 1993; Kirwan & Stark, 2007; Naveh-Benjamin, 2000; Yonelinas, 2002; Yassa and Stark, 2011). Usually, older adults are overconfident than younger adults for their FOK predictions, even if the items are unrecalled at the moment, implying aging may reduce accurate calibration of future memory. Previous research, as mentioned before, has

shown relying on context in general not only enhances FOK predictions but also helps older adults to enhance their memory performance, as context can provide a supportive aid in recalling partial information for target. If older adults take context into account at study phase to increase their memory performance for target, in a subsequent cued recall phase, we can not only expect them to recall more of target information in a context-pair association they have seen before but also feel more confident, as a result, higher FOK ratings for such recalled items with old context.

Experiment 1

Overall, from the literature review, we know in general, context can influence our main interest, FOK judgments in older adults. Context also influences memory performance in older adults. Additionally, although context can act as distractors, research in general also shows context usually provides a supportive memory aid, especially for older adults.

The purpose of Experiment 1 was to understand the underlying process of using background information on memory and metamemory beliefs. Our first aim was to investigate whether people implicitly associate unrelated background information with FOK predictions and actual target memory. In this case of implicitness, we did not ask any participant to ignore context. Rather we provided a context condition to see if they could recall them and asked a post-test questionnaire on whether background information helped them to inform their predictions and memory. Our second aim was to investigate whether people treat all backgrounds the same or whether some background information informs FOK predictions and target recall better than others. In this case, we provided three different background contexts: original context, repaired context, and novel context conditions. Finally, our third aim was to investigate whether the rate

of using such a background to inform FOK predictions and memory is differentially influenced by age, younger and older adults.

Experiment 1 Main Question: Do FOK predictions differ as a function of age group across three different context conditions (i.e., an age x condition interaction), in the original context, repaired context, and novel context?

Hypothesis:

FOK: Expecting an age X context interaction for FOK unrecalled items where younger adults will have higher FOK ratings in original context=repaired context>novel context condition whereas older adults will have the highest rating for original context>repaired= novel context (no significant difference).

The main effect of age in unrecalled items: In unrecalled items, older adults may have higher FOK ratings compared to younger adults. No significant difference in recalled items.

Recall: Expecting an age X context interaction such that younger adults may have a higher recall rate for original=repaired>novel context. For older adults, the cued-recall may be highest for the original>repaired=novel context.

Reason: For all participants, if they incorporate context into their FOK predictions, original and repaired context have more chance of reminding target (if they associated target-background at study phase) than a novel context. This effect may be shown in younger adults. In older adults, we must consider two age-related reduction in cognition: 1) more reliance on context cues than younger adults due to distracted attention and 2) poor binding ability that influences cognitive performance as we age. So, for the original context in older adults, we have the same reasoning as in younger adults. But in a repaired context, it is more likely older adults may confuse the correct target associated with the correct background when rearranged. Hence,

an original context might provide maximum target information compared to a novel or repaired context as shown in the recall, but this confusion may make older adults give similar high FOK ratings in repaired context as original context.

2. METHODOLOGY EXPERIMENT 1

Participants

Power Analysis. Using the G*power software program (Faul, Erdfelder, Lang, & Buchner, 2007), a power analysis revealed that Experiment 1 would require 60 participants (30 younger and 30 older adults) and Experiment 2 would require 60 participants (30 younger and 30 older adults). An alpha level of .0167 would be needed to detect a medium (f=.25) effect (Doss, Picart, & Gallo, 2018) with a power of 0.80 using a 2 (Age: Younger and Older adults) x 3 (Context Conditions: Original, Repaired and Novel Context) mixed design repeated measures ANOVA with context conditions as within-subject and age as a between-subject factor.

Recruitment. Thirty younger adults of age range 18-25 years were acquired from PY 101 Subject Pool from the University of Alabama. Thirty older adults of age range 60 years or above in person were acquired from Tuscaloosa Community. Thirty-three older adults of age range 60 years or above were recruited from Amazon Mechanical Turk (M Turk). Four participants from M Turk were screened out as they scored a high score of GDS=14 indicating depression. Thus, 29 older adults' data from M Turk were included in the analysis. Overall, 30 younger adults and 59 older adults were included in the analysis. Younger participants were compensated with course credits and older participants in person were rewarded $10 and M Turk was awarded $5 for participation.

Stimuli

For context and cue-target photographs, 160 background scenery and 160 pictures of everyday objects were acquired, respectively. Additional 160 related pictures of everyday objects were acquired for similar lure pairs. All pictures were acquired from the Eakin database (Eakin Memory and Metamemory Lab MSU), ImageNet database (www.image-net.org), and the Internet.

Pair and context relatedness for stimuli. First, unrelated pictures were paired randomly creating sixty unrelated picture pairs in total. These pairs were then shown to 50 younger adult participants to obtain average pair-relatedness ratings from UA Subject Pool. Participants viewed 40 unrelated picture pairs. Under each labeled pair, a 5-point scale was provided, and participants were asked to rate how visually related the pictures in the pair. We created similar lures to use the lures in recognition memory. Each cue-target pair was superimposed with an unrelated context photograph in the background. The purpose of the test was to ensure that the pairs and contexts were unrelated. Under each pair, a 5-point scale was provided. Participants were asked to rate the relatedness of the unrelated context and overall pairs (1=least unrelated to 5=most unrelated). A One-Way ANOVA was conducted to ensure the 12 list do not statistically differ significantly from each other in terms of cue-target relatedness, F (11,132) = 0.74, p = .70 (M = 1.15, SD = 0.53), cue-lure relatedness, F (11,132) = 1.40, p = 0.18, (M = 1.11, SD = 0.53), target-lure relatedness, F (11, 132) = 0.14, p = 1.00, (M = 2.78, SD = 0.76), original context relatedness, F (11, 132) = 0.67, p = 0.76, (M = 0.99, SD = 0.37), novel context relatedness, F (11, 132) = 0.94, p = 0.50, (M = 1.06, SD = 0.47). The descriptive statistics show relatedness ratings are low and do not differ between lists.

Counterbalances. Using the ratings above, we created 12 counterbalances. Each counterbalance had the 12-block list counterbalanced. Within each counterbalance, there were two blocks of study-test-FOK phase consisting of 18 study-cued recall-FOK items in Block 1 and 18 study-cued recall-FOK items in Block 2, making a total of 36 cue-target pairs to study. In each block of cued recall, 18 items were divided equally to pair with context conditions: 3 items were paired with original contexts, 3 items were paired with novel context and 3 items were paired with a repaired context. In each block of the old/new and recollection/familiarity phase, 18 items were divided equally to pair as 9 cue-original targets and 9 cue-lure-target.

Each counterbalance also had 18 blocks of old contexts (all three context types in study and test phase) and 18 blocks of the new context. Within each block of old/new and recollection/familiarity phase, 18 contexts were divided equally to pair as 9 original contexts and 9 novel contexts. To avoid confounding "oldness" variable (whether old if they have seen it before in the study vs. test phase) the new context in this phase were background pictures that were never shown before as to represent an actual novel context. Repaired context can only be created with associated pairs and cannot be created with the only context and thus, was not shown.

Materials

Informed Consent and Debrief

All participants received and signed informed consent for a protocol approved by the Institutional Review Board (IRB) protocol. Debrief was not necessary as there was no deception as approved by the IRB protocol. Younger participants completed the study online. Older

participants from the community conducted the study in person. Older adults recruited from M Turk conducted the study online.

Demographics

See Table 2 for statistics. Based on self-report, all participants had normal to corrected vision. They were asked about their age, race, SES, and education. All participants were below the cut-off value and healthy.

Global Cognition

Older adults in person were given St. Louis University Mental Status (SLUMS) questionnaire (Tariq, Tumosa, Chibnall, Perry, & Morley, 2006) to measure orientation, short-delay verbal memory, visuospatial construction and executive function in older adults, reported in Table 2. SLUMS is a 30-pts questionnaire with 11 questions in total. Questions 1-3 test for attention, recall and orientation, questions 4 and 7 test for delayed recall and interference, questions 5 and 6 test for numeric calculations and registration, question 6 tests for memory: immediate recall with interference (time constraint), question 8 tests for registration and digit-span, question 9 tests for visual span, question 10 tests for visuospatial and executive function and question 11 tests for executive function and extrapolation. Since the current study requires the use of memory and executive function in older adults, SLUMS is a more comprehensive questionnaire than MMSE for our study.

Unlike MMSE (Folstein, Folstein, & McHugh, 1975), SLUMS is a free available questionnaire. In contrast to MMSE, SLUMS is more sensitive in detecting changes in cognition for a couple of reasons including being sensitive in demographic variables such as the difference in education, additional cognitive domains such as logical memory, size differentiation, verbal

fluency, is more sensitive to changes in executive function (the animal naming and clock drawing items) and has less skew and less mean than MMSE in a normal non-veteran population signifying less ceiling effects (Howland, Tutsouka, Smythe, & Sajatovic, 2017). Older adults should score with a range of 27-30 (high school education and above) or 25-30 (education less than high school) to rule out cognitive decline.

Older adults in M Turk were given the Short Portable Mental Status Questionnaire (SPMSQ) (Pfeiffer, 1975) to screen for dementia or cognitive deficiency. This is a 10 item questionnaire and participants must score above 8 to rule out cognitive decline. We modified three questions so as not to violate the M Turk confidentiality rule (see Appendix B for the SPMSQ questionnaire modified for M Turk approved by IRB and Table 2 for scores). The final questionnaire had 8 questions and participants had to score all correct for all 8 questions. For age verification, M Turk has the option to add in the categories we want and select the age range. These links with their worker's account and appropriate participants are then selected. For age accuracy in M Turk, we have our redundancy check embedded in the study. In the Basic Demographics pre-questionnaire, we asked age in numbers ("How old are you?"), while in the SPMSQ, we ask age in the year ("Were you born before 1964?") to verify. Additionally, in recent years, there are changes to workers account such that there are tax implications, completion rate, and prohibition of duplicate account. All these factors together combine to verify a good worker with an eligible age range. This is similar to age verification to older adults in person, as we also don't verify age with passport or driver's license. All older adults included passed the SLUMS and SPMSQ tests, depending on the sample (see Table 2).

Depression Scale

Older adults were given the 15-item short Geriatric Depression Living Scale (GDS) (Yesavage et al 1982; Sheikh & Yesavage, 1986) reported in Table 2. GDS is a self-report questionnaire that screens out depression in older adults that may lead to memory complaints (Pearman & Storandt, 2004). This shortened form of the scale has been shown to a valid and reliable source to screen for depression (Sheikh & Yesavage, 1986). Scores range from 0-15 with higher scores indicating depression. Participants should score <5 to rule out depression. All older adults included did not have depression (Table 2).

Self-Esteem Scale

Older adults were given the 10-item Rosenberg Self Esteem Scale (Rosenberg, 1965) to measure their self-esteem (both positive and negative about self) which also is a predictor for memory performance in older adults (Pearman & Storandt, 2004), reported in Table 2. All items are scored on a 4-point scale from Strongly Agree to Strongly Disagree. Scored are added for all 10 items. Younger adults also completed the questionnaire. Scores range from 0-30 with 0-15 low self-esteem, 15-25 normal self-esteem, and 25-30 high self-esteem. Participants should score >=14 to rule out low self-esteem. All older adults included had moderate to high self-esteem (see Table 2).

Motivation

Motivation can also play a role in memory performance in participants. Participants with high cognition are more motivated towards the task and are less influenced by misleading cues, more likely to reflect upon feedback about task performance (Anseel, Lievens, & Schollaert, 2009) and more accurately predict future performances, such as metamemory and metacognition (Reinhard & Dickhauser, 2009, Cacioppo, Petty, Kao, & Rodriguez, 1986), process information

at a conceptual level and are confident in their judgments compared to those with low cognition (Haugtvedt, Petty & Cacioppo, 1992). All participants were given the 18-item Need for Cognition Scale (NCS) (Cacioppo, Petty, & Kao, 1984) to measure their intrinsic motivation to enjoy and engage in cognitively challenging activities. All older adults included had moderate to high motivation (see Table 2).

IQ and Education Scale

As an additional measure of IQ and education, younger and older adults were asked to complete the ETS Vocabulary questionnaire to integrate into the online computerized study, reported in Table 2. Higher scores indicate higher achievement. In Experiment 2 we also added a reading skill assessment for older adults as an added measure as IRB protocol revision was taking very long so we could not use it for Experiment 1. This difference did not influence study participation first, it was a post-test questionnaire, second, all older adults completed the ETS vocabulary and had high scores, and third, they were all highly educated (see Table 2). They were asked to complete the Wide Range Achievement Test (WRAT-4) (Wilkinson, 1993), reported in Table 2. The test is used to assess the basic academic skills of participants in reading. It can be used for ages 5-90 years. The reading subtest assesses an individual's ability to recognize and name 15 letters and pronounce 55 words out of context. Scoring of the reading test is not affected by unusual pronunciations due to regional accents or speech difficulties. Scores range from 45-70 with higher scores indicating higher achievement.

Procedure

See Figure 1 for the study procedure.

Study Phase

For each block of items, participants viewed 18 unrelated cue-target labeled picture pairs (e.g. mop-airplane) superimposed on a black and white context photograph for 6 seconds with a 1-second inter-stimulus interval (Gutchess & Park, 2009). Participants were instructed to study the pairs for a future memory test. Picture pairs were in color and were presented in a white box, which was superimposed on black and white context scenery.

Cued Recall and FOK judgment phase

The study phase was followed by a cued recall phase for every 18 items in which the original cues with a missing blank for the target (*mop-_?_*) were presented with context manipulations for 3 items each. Context manipulation is given in this phase as separate data is needed for FOK judgments with context manipulations for FOK analysis. Cued recall for each item was followed by a FOK judgment for every 18 items where participants provided a FOK rating. Both cued recall and FOK judgment phase were untimed, per item by item basis, and participants had to respond before moving on to the old/new recognition phase.

Old/New Recognition Judgment and Recollection/Familiarity Ratings for pairs

In this phase, 18 cue-target pairs were shown without context (Gutchess & Park, 2009). The procedure for the current phase has been adapted and modified for the current study by Gutchess et al. (2007). Generally, past studies on contexts at recognition memory had both old and new (lures) stimuli superimposed on context conditions (Gutchess et al., 2007; Hanczakowski et al., 2017), but with explicit instructions to ignore the context. Both tasks, adding new stimuli, and instructions to ignore context, control for source memory errors. In the current study, context is only manipulated in cued recall and not manipulated in recognition

phase as this may lead participants to make source memory error, generally in older adults (Cansino et al., 2013; Chalfonte & Johnson, 1996; Glinsky, Rubin, & Davison, 2001) in the old/new recognition phase to discriminate if the novel or re-paired context were from the cued recall or the old/new recognition judgments.

In the old/new judgments, participants responded whether the pairs shown are "old" or "new". For each trial, the participant viewed 18 number of items. 9 items were original cues paired with original target whereas 9 items were original cues paired with lure targets (a red pen instead of the original blue pen studied). The original studied items that are shown as lures will not be presented at recognition to avoid recall-to reject strategies by participants. For example, participants might see an old pair first and judge it as "old" but they may see similar lure next and judge it as "new" not because they could distinguish between "old" or "new" pairs but because they might base their judgment on the previous old pairs. During the test, instruction what is an old vs. new response was given to participants: "*Is this an old pair or a new pair? Respond to OLD if you think you studied this exact intact pair before. Respond NEW otherwise.*"

After responding whether they think the item is old or new, for each item, participants then provided recollection/familiarity ratings on their memory for that particular item. Instructions on what is recollection and familiarity were provided during the study which was adapted from McDonough and Gallo (2008) and Gallo, Meadow, Johnson, and Foster (2008) as shown below. Old/New and Recollection/Familiarity phase were untimed, item by item basis, and participants were required to respond by moving on to the next item.

> *"Actually Recollect: You may have a very strong or vivid memory of the picture pairs because you visually remember the actual picture pair or what you were thinking when*

the pair was presented. Please note these ratings are not about one's confidence - you might be very confident that the item was studied because you recollect one detail or many detail

Drag the slider below from 1-7 to rate the strength of your recollection.

A rating of 1 would indicate poor and vague recollection. A rating of 7 indicate that recollection is very strong, or vivid.

Very Familiar: You may have a weak or vague recollection of the pairs if you think the pairs were presented but you cannot actually recollect any details about the object or the task it was presented in. You "just know" it was studied. Please note the ratings are not about one's confidence - you might be very confident that an item was studied (because it is very familiar), but you cannot recollect any details about the presentation of the object.

Drag the slider below from 1-7 to rate the strength of your familiarity.

A rating of 1 indicates the pair of pictures have a very low feeling of oldness. A rating of 7 indicates the pair feels very old"

Old/New Recognition Judgment and Familiarity/Recollection Ratings for Context

A second old/new recognition judgment with recollection/familiarity phase was conducted similar to the previous one with the exception that only the context manipulations were shown without the pairs. There were also 2 blocks as the previous phase. This additional phase was added to re-affirm on whether context alone would influence FOK ratings and helps us answer whether participants paid attention to context when making their FOK ratings or for their memory performance. For each block participants viewed 36 contexts with 18 contexts that

were seen with items on all three context manipulations and 18 contexts never seen before with items. These novel contexts were not shown before in any form during the previous phases for items. These contexts were used to provide additional buffers to avoid source memory confusion on whether the novelty of the context is based on the study phase or between the cued recall phases. We also could not show the context manipulation only for re-paired context since it would need to have a focal item superimposed on the context. In other words, if a participant is saying "Old" to a context that was shown at cued recall in an Original Context, Repaired Context, or Novel Context**,** it would be a hit because regardless of the context condition, they have seen it at cued-recall phase for the pairs. If a participant is saying "Old" to the novel context exclusive to the context only condition phase, then it is a false alarm because they have not seen this context before either at cued recall for pairs or at context condition only.

After the cued recall and FOK judgment phase, both younger and older adults will complete several questionnaires (ETS Vocabulary, Rosenberg Self-Esteem, Need for Cognition scales, GDS for older adults only).

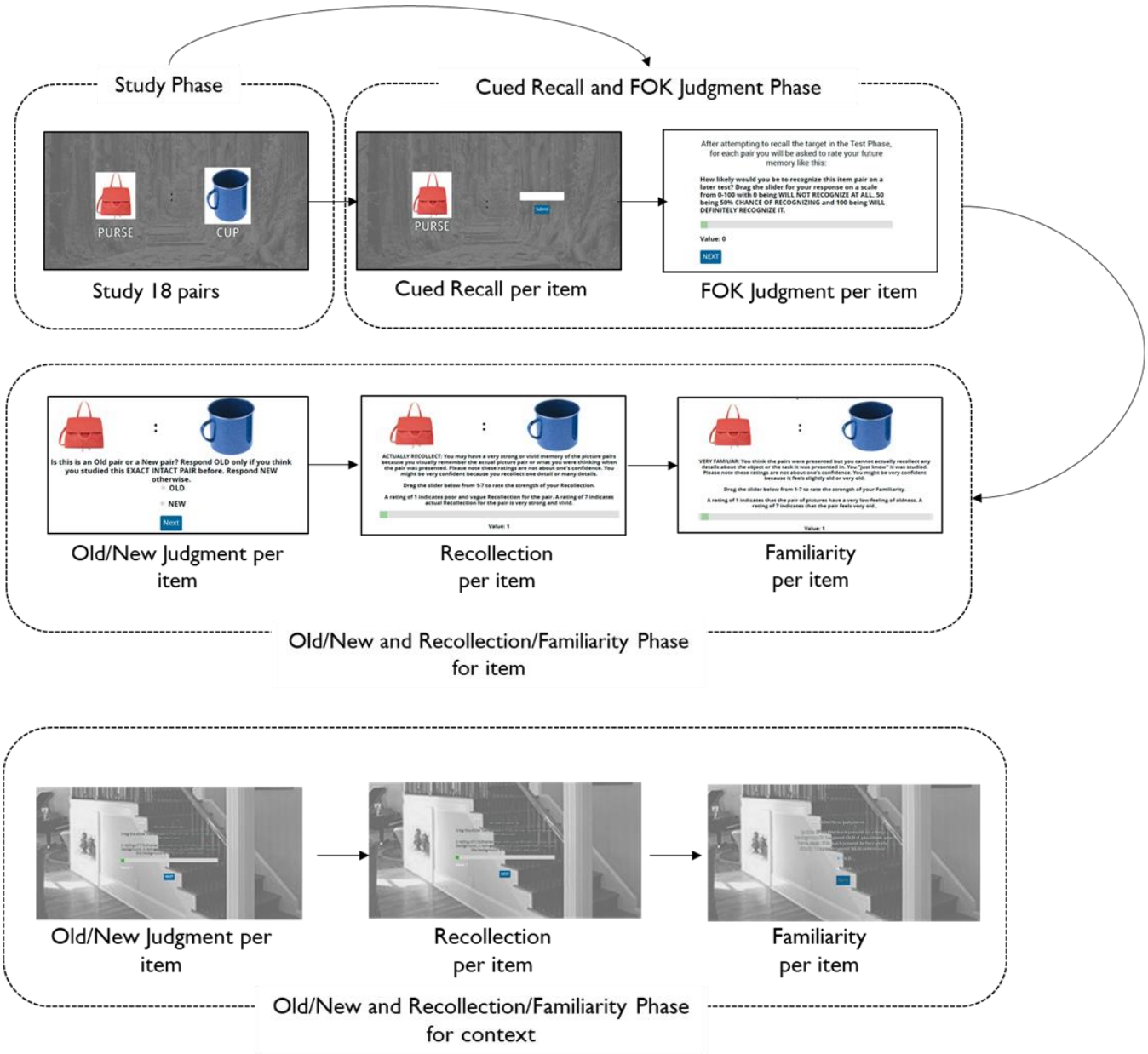


Figure 1: Study Procedure
FOK: Feeling of Knowing Judgment

3. RESULTS EXPERIMENT 1

For each analysis, if there is a 3-way interaction, then it will be given priority and reported in detail while significant main effects or 2-way interaction will be mentioned briefly. In absence of 3-way interaction, main effects and 2-way interaction will be reported. This is done to avoid confusion with multiple analyses. FOK scale has been adjusted to 70 for both experiments so that effects can be seen clearly.

Covariate Analysis for Age

A bivariate correlation revealed most covariates correlated with one another (Table 3).

Table 3

Correlations of Covariates for Experiment 1 by Age

		ED_ TOTAL	FATHER ED	MOTHER ED	INCOME	VERBAL SCORE	SELF ESTEEM	NCS	SES
ED_TOTAL	Pearson Correlation	1.00	-.14	.24*	-.04	.27*	.13	.15	.32**
	Sig. (2-tailed)		.19	.02	.69	.01	.22	.16	.00
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00
FATHERED	Pearson Correlation	-.14	1.00	.08	.18	-.06	-.07	-.02	.67**
	Sig. (2-tailed)	.19		.46	.09	.57	.54	.83	.00
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00
MOTHERED	Pearson Correlation	.24*	.08	1.00	-.12	.18	-.10	.03	.66**
	Sig. (2-tailed)	.02	.46		.26	.10	.37	.75	.00
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00
INCOME	Pearson Correlation	-.04	.18	-.12	1.00	-.29**	-.07	.14	.36**
	Sig. (2-tailed)	.69	.09	.26		.01	.53	.18	.00
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00
VERBAL SCORE	Pearson Correlation	.27*	-.06	.18	-.29**	1.00	.00	.23*	.04
	Sig. (2-tailed)	.01	.57	.10	.01		.99	.03	.69
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00
SELF ESTEEM	Pearson Correlation	.13	-.07	-.10	-.07	.00	1.00	-.01	-.08
	Sig. (2-tailed)	.22	.54	.37	.53	.99		.94	.44
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00
NCS	Pearson Correlation	.15	-.02	.03	.14	.23*	-.01	1.00	.10
	Sig. (2-tailed)	.16	.83	.75	.18	.03	.94		.37
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00

SES	Pearson Correlation	.32**	.67**	.66**	.36**	.04	-.08	.10	1.00
	Sig. (2-tailed)	.00	.00	.00	.00	.69	.44	.37	
	N	89.00	89.00	89.00	89.00	89.00	89.00	89.00	89.00

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

An independent sample with Age as an IV and covariates as DVs were conducted to see which covariates differed across the two age groups (young vs. old) and help in deciding to choose the most important covariates for the analysis. We found across age, 4 covariates were significant: Younger adults had a lower level of total education, t (87) = 3.29, p<0.00, had a higher level of father's education t (87) = 3.02, p <0.00, had a higher level of income t (87) = 5.29, p <0.00 (which we suspect was parent's income as they are dependents) and lower level of verbal score t (87) = 7.00, p <0.00, compared to older adults. Due to high collinearity among some covariates such as verbal score with income and total education, we created a composite score of the significant covariates by combining their z-scores. This composite score was then entered for each analysis when we found an age effect without the covariates.

Analysis by Age (Younger adults vs. All Older Adults)

Feeling of Knowing (FOK) Mean Ratings by Age

We wanted to investigate whether metamemory, in other words, FOK mean ratings differed in terms of age and context conditions. To understand this question, we conducted a 2 (Response: Recall vs. Unrecall) x 2 (Age: Young vs. Old) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with response and context as within-subject and age as a between-subject factor (Fig 2).

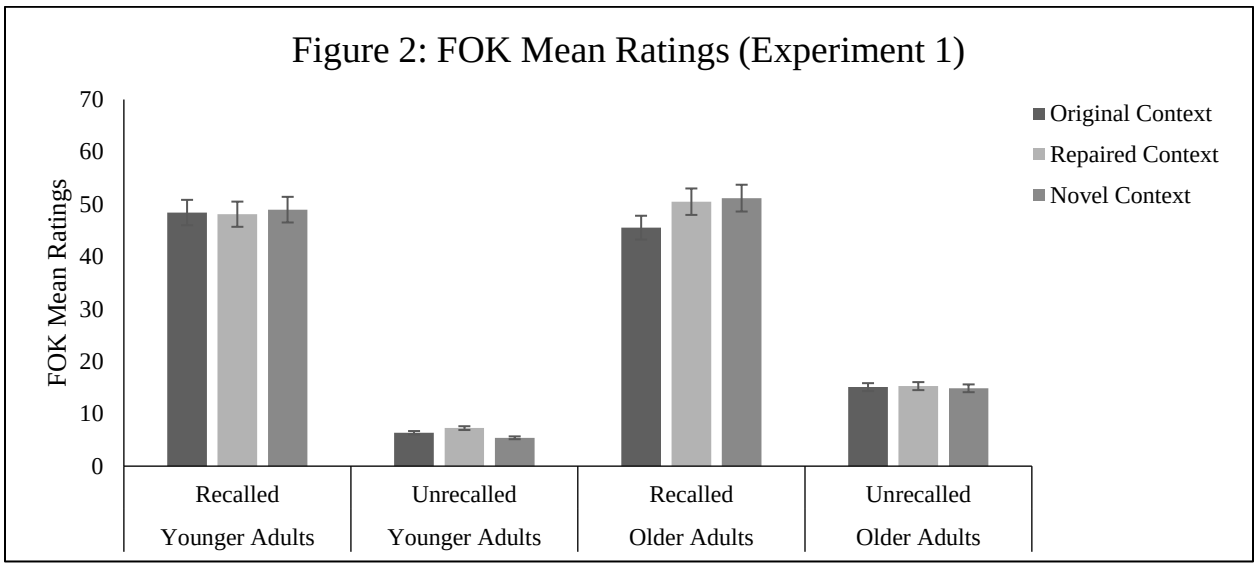


Figure 2: Mean Feeling of Knowing (FOK) ratings between recalled and unrecalled items across context and age. Scale range from 0-100. There was only a main effect of response type, F (1,87) = 131.42, p< 0.00. No significant interaction of age or context were found.

Main effects: Results revealed a main effect of response type, F (1,87) = 131.42, p< 0.00 such that participants gave higher FOK ratings for recalled items (M = 48.54, SD = 28.22) compared to unrecalled pairs (M = 12.14, SD = 12.90), t (88) =11.74, p< 0.00.

Interaction: No other significant interactions were found, context by age, F (2, 174) = 1.26, p = 0.29, response type by age F (1, 87) = 1.19, p = 0.28, context by response type, F (2, 174) = 0.26, p = 0.77, context by response by age, F (2, 174) = 0.73, p = 0.48.

Overall results support hypothesis 3 with no interaction effects of age or context on FOK ratings.

Cued Recall by Age (Experiment 1)

We wanted to investigate whether objective memory such as cued recall would differ with age and context conditions. To understand this question, we conducted a 2 (Age: Young vs.

Old) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with age as a between-subject factor and context as a within-subject factor.

Main effects: No significant main effects were found for context conditions, F (2, 174) = 0.24, p = 0.79 (Table 4).

Table 4

Descriptive Statistics for Cued Recall Experiment 1 by Age

	Age	Mean	Std. Deviation	N
Recall_original	YA	.43	.31	30.00
	OA	.32	.26	59.00
	Total	.36	.28	89.00
Recall_repaired	YA	.41	.31	30.00
	OA	.32	.26	59.00
	Total	.35	.28	89.00
Recall_novel	YA	.44	.33	30.00
	OA	.31	.26	59.00
	Total	.36	.29	89.00

Interaction: No significant interactions were found, context by age, F (2, 174) = 0.42, p = 0.79.

Overall results do not support any of the hypotheses above. Results suggest older and younger adults are equivalent when making cued recall attempts and context condition has no effects when making such attempts.

Old/New Judgment by Age

We also measured another kind of objective memory besides cued recall. This objective memory was old/new judgment recognition memory and was measured after participants

completed their cued-recall-FOK phase for all pairs (see Fig 1 for Study Procedure). These judgments measure a participant's ability to discriminate between target pairs from lure pairs that were perceptually similar to target pairs. The purpose was to explore whether participants' age played a role in recognizing target pairs better compared to lure pairs across context conditions in the cued recall phase.

Thus, two measures were obtained from old/new judgment: hit and false alarms. In the picture pair phase, hit rates were calculated as the proportion of correctly selecting "old" for the original cue-original target pair. In context only condition phase, hit rates were calculated as the proportion of correctly selecting "old" for original, repaired, and novel contexts which they have seen at cued recall. In the picture pair phase, false alarms were calculated as the proportion of incorrectly selecting "old" for the cue-lure target pair. In context only condition phase, a false alarm was calculated as the proportion of incorrectly selecting "old" for the new novel context exclusive at the context only condition, which participants have never seen before at cued recall.

After each measure was collected, for the picture pair phase, we calculation a discrimination score (Hits-FA) for each context condition—original, repaired, and novel contexts, reported below. We did not calculate discrimination score for context condition only but reported hits and false alarms separately since hits and false alarms did not have the same context conditions due to presentation of context scenes without picture pairs at this stage.

Discrimination Score (Hits-False Alarms) for Picture Pairs by Age. A 2 (Age: Young vs. Old) x 2 Context at Cued-Recall Test (Original, Repaired, Novel) mixed ANOVA with age as a between-subject factor and context as a within-subject factor was conducted to test the effect of

context and age on participant's discrimination ability between target vs. lure pairs across age and context conditions.

Main effects: No significant effects were found for context conditions, F (2, 174) = 0.60, p = 0.56 (Table 5).

Table 5

Descriptive Statistics for Discrimination Scores Experiment1 by Age

	Age	Mean	Std. Deviation	N
Pairs_Original	YA	.45	.35	30.00
	OA	.35	.41	59.00
	Total	.38	.39	89.00
Pairs_Repaired	YA	.41	.35	30.00
	OA	.30	.25	59.00
	Total	.33	.29	89.00
Pairs_Novel	YA	.40	.35	30.00
	OA	.34	.30	59.00
	Total	.36	.32	89.00

Interaction: No significant effects were found, context by age, F (2, 174) = 0.23, p = 0.79.

Overall results from hits and false alarms from picture pairs seem to suggest background type or age does not influence memory discrimination for picture pairs or when participants make errors in their memory discrimination. In other words, we see an age equivalence for memory discrimination ability in younger and older adults without any influence of background image.

Hit for Context Condition Only by Age. A 2 (Age: Young vs. Old) x 3 Context at Cued-Recall Test (Original, Repaired, Novel) mixed ANOVA with age as a between-subject

factor and context as a within-subject factor was conducted to test the effect of context and age on hit rates of background contexts only.

Main effects: Result revealed a significant main effect of context, F (2,174) =7.16, p = 0.01. A paired sample t-test revealed only two effects. Overall participants have higher hit rates for original context (M = .46, SD = 0.31) compared to novel context (M = 0.37, SD = 0.29), t (88) =3.18, p <0.00 Also, participants have higher hit rates for repaired context (M = 0.45, SD = 0.28) compared to novel context, t (88) = 2.41, p = 0.02. Ratings between original and repaired context was not significant, t (88) = 3.18, p = 0.48.

There was also a main effect of age for between-subject test F (1,87) =9.96, p = 0.002. An independent sample t-test revealed overall for context conditions, hit rates was significantly higher for younger adults (M = 0.54, SD = 0.21) compared to older adults (M = 0.37, SD = 0.24), t (87) = 3.16, p = 0.002.

Interaction: No significant interactions were found, context by age, F (2, 174) = 2.05, p = 0.13.

Covariates: After controlling for covariates, we found the age effect to be still significant F (1, 86) = 8.25, p<0.00.

False Alarm for Context Condition Only by Age. An independent t-test was conducted with age as a between-group as independent variable and ratings of false alarms for novel context exclusive to the context condition as the dependent variable. These novel contexts were exclusive in the context conditions only, meaning participants have not seen these novel contexts before at the study or test phase. We wanted to see the false alarm rate in this context only

condition-stage (saying “old” to a context never shown before, which in this case would be a lure context scene).

Effects: No significant effects were found, t (87) = 1.47, p =0.15 (Table 6).

Table 6

Group Statistics for False Alarm for Context Condition Stage Experiment 1 by Age

	Age	N	Mean	Std. Deviation	Std. Error Mean
Context_FA	YA	30.00	.40	.18	.03
	OA	59.00	.33	.22	.02

Interaction: Not applicable

Overall results suggest participants of all ages take context into account, especially for old scenes compared to new scenes. In general, supporting previous research, younger adults have higher hit rates compared to older adults for context only condition. Results may hint that even younger adults take some form of context into account when studying the pairs to correctly recognize and discriminate them from novel scenes.

Recollection for Picture Pairs by Age

Recollection for Items (Target vs. Lures) for Picture Pairs. We wanted to investigate whether subjective ratings of the number of details recollected differed in terms of age and context conditions for target pairs. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Pair: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and pair as a within-subject and age as a between-subject factor to investigate the effect of context conditions and age on recollection ratings for the pairs. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted

to test whether recollection differs depending on whether it was a target or a lure, in addition to context and age conditions.

Main effects: No significant effects were found for context, F (2, 174) = 0.68, p = 0.51 or item, F (1, 87) = 0.42, p= 0.52 (Table 7).

Table 7

Descriptive Statistics for Recollection Ratings for Item ((Target vs. Lure) Experiment 1 by Age

	Age	Mean	Std. Deviation	N
Recollection_Target_Original	YA	2.6	.98	30.00
	OA	2.9	1.05	59.00
	Total	2.85	1.03	89.00
Recollection_Lure_Original	YA	2.51	1.01	30.00
	OA	3.04	1.43	59.00
	Total	2.86	1.32	89.00
Recollection_Target_Repaired	YA	2.72	1.08	30.00
	OA	3.09	.94	59.00
	Total	2.96	1.00	89.00
Recollection_Lure_Repaired	YA	2.50	1.19	30.00
	OA	3.19	1.13	59.00
	Total	2.96	1.19	89.00
Recollection_Target_Novel	YA	2.66	1.19	30.00
	OA	3.17	.987	59.00
	Total	3.00	1.08	89.00
Recollection_Lure_Novel	YA	2.41	1.19	30.00
	OA	3.23	1.18	59.00
	Total	2.95	1.24	89.00

Interaction: No significant effects were found, context by age, F (2, 174) = 0.99, p = 0.37, item type by age F (1, 87) = 2.62, p = 0.11, context by item type, F (2, 174) = 0.10, p = 0.91, context by item by age, F (2, 174) = 0.05, p = 0.95.

Overall, results suggest context type or age does not influence subjective ratings for the recollection of intact picture pairs or lure picture pairs. In other words, the result suggests an age equivalence for subjective memory, specifically recollection ratings for picture pairs without any influence of background image.

Recollection for Responses (Old vs. New) for Picture Pairs by Age. We also wanted to investigate whether subjective ratings of the number of details recollected differed in terms of age and context conditions for items participants deemed as "Old", in other words, pairs they think they have seen before. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Response: Old vs. New) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor to investigate the effect of context conditions and age on recollected ratings for the pair responses given during their old/new judgment phase. This analysis considered whether the targets or lures were correctly endorsed as "Old" or "New."

Main effects: There was a main effect of response, F (1, 87) = 268.72, $p < 0.00$ and a paired sample t-test showed old responses (M = 4.42, SD = 1.46) had higher ratings than new responses (M = 1.44, SD = 0.87), t (88) = 17.59, p <0.00.

Interaction: There was a significant 2-way interaction of response by age, F (1, 87) = 9.07, p <0.00 followed by a significant 3-way interaction between context by response by age F (2,174) = 4.25, p = 0.02 (Fig 3).

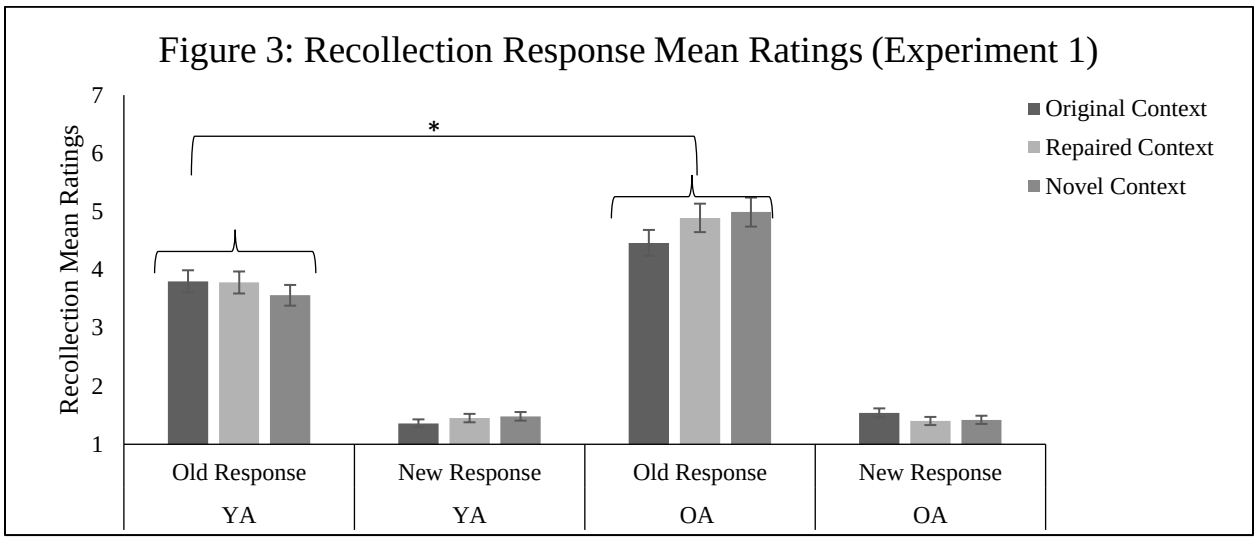


Figure 3: Mean recollection ratings for pairs between old response and new response across context and age. Scale range from 1-7. There was a significant 3-way interaction of context by age by response, F (2,174) = 4.25, p = 0.02.

To understand a 3-way interaction, we first did two 3 (Context: Old, Repaired, Novel) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and age a between measure on Old items and New items as the dependent variable for each test respectively.

For old items, we found a 2-way response by age interaction, F (1,87) = 9.06, p <0.00. To understand this interaction, first we did an independent t-test and results revealed three significant effects. Across old responses, older adults (M = 4.46, SD = 1.54) had marginally significant higher ratings for pairs compared to younger adults (M = 3.80, SD = 1.49) in original context, t (87) = 1.93, p = 0.06. Older adults (M = 4.89, SD = 1.44) had significant higher ratings for pairs compared to younger adults (M = 3.78, SD = 1.72) in repaired context, t (87) = 3.20, p< 0.00. Older adults (M = 4.99, SD = 1.46) also had significant higher ratings for pairs compared to younger adults (M = 3.60, SD = 1.58) in novel context, t (87) = 4.14, p<0.00.

Then we split the file into young and old age groups. We re-ran the ANOVA testing for simple effects. For younger adults, no significant difference in context conditions was found. For older adults, there was a significant main effect of context conditions for old responses, F (2, 116) = 8.69, $p < 0.00$. A follow up paired sample t-test within older adults revealed only one effect among context conditions. For old responses, there was a higher recollection rating for picture pairs in novel context (M = 3.20, SD = 0.90) compared to original contexts (M = 3.00, SD = 1.06), t (58) = 3.32, p <0.04. For new items or within younger adults there were no significant effects.

Covariates: The age effect was still significant after controlling for covariates, F (1, 86) = 7.12, p<0.00.

Overall, the results show age and context conditions only matter for picture pairs which were deemed as "Old" in the old/new judgment phase. Older adults seem to consider the context when making their subjective recollection ratings for picture pairs. Interestingly, their recollection for pairs labeled as "old" seemed to be higher for novel context compared to an original context.

Familiarity for Picture Pairs by Age

Familiarity for Items (Intact Target vs. Lures) for Picture Pairs by Age. We wanted to investigate whether subjective ratings of a feeling of "oldness" differed in terms of age and context conditions for target pairs. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and item as a within-subject and age as a between-subject factor to investigate the effect of context conditions and age familiarity ratings of the item type. This

analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether familiarity differs depending on whether it was a target or a lure, in addition to context and age conditions.

Mauchly's test revealed assumption of sphericity has been violated for context, χ^2 (2) = 7.39, ε = 0.92 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.92>0.75) *p* = 0.03 and for context by item χ^2 (2) = 15.53, *ε* = 0.86 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.86>0.75), *p* <0.00, therefore degrees of freedom were corrected using Huyunh-Feldt estimates of sphericity.

Main effects: Results revealed a main effect of context conditions, *F* (1.91, 165.98) = 15.04, *p* <0.00. A paired sample t-test revealed picture pairs in original context (*M* = 2.99, *SD* = 0.94) had higher ratings compared to novel contexts (*M* = 2.76, *SD* = 0.94), *t* (88) = 2.30, and *p* < 0.03 and repaired context (*M* = 3.15, *SD* = 0.11) had higher ratings compared to items in novel context *t* (88) = 3.71, *p* < 0.00, respectively.

There was also a main effect of item, *F* (1,87) = 8.18, *p* = 0.005. A paired sample t-test revealed intact target items (*M* = 4.12, *SD* = 1.17) having significantly higher familiarity rating than lure pairs (*M* = 2.71, *SD* = 1.11), *t* (88) = 12.41, *p* <0.00.

Interaction: Main effects were followed by a significant 2-way interaction of context by item for picture pairs *F* (1.78, 153.78) = 11.66, *p* < 0.00 (Fig 4).

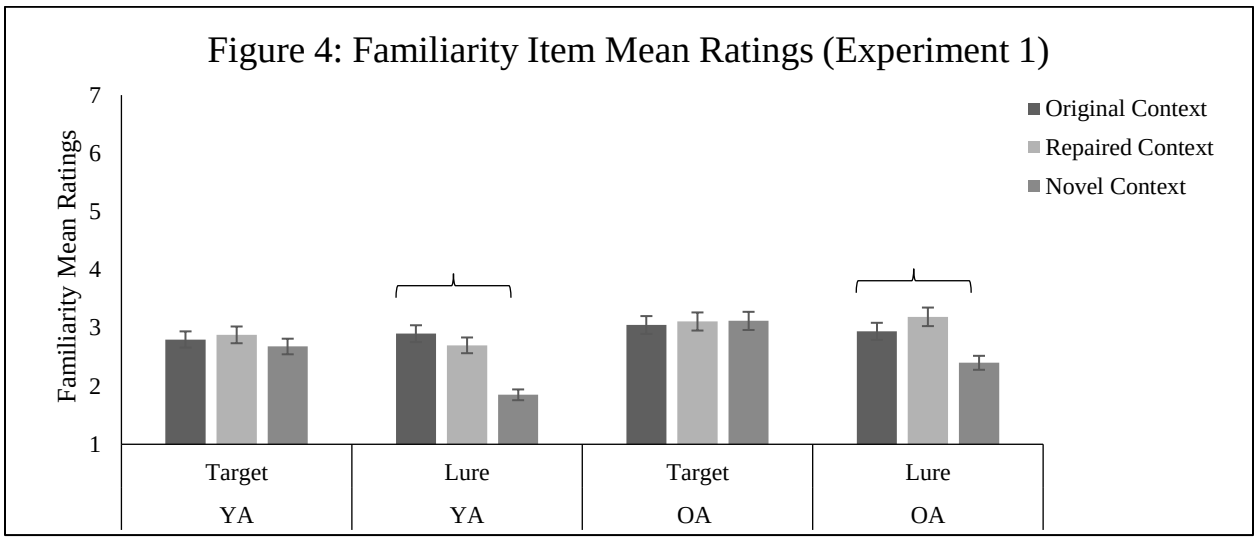


Figure 4: Mean familiarity ratings for pairs between target and lures across context and age. Scale range from 1-7. There was a significant 2-way interaction of context by item across lures, F (1.78, 153.78) = 11.66, p < 0.00

To understand this 2-way interaction, a plot revealed the differences could be due to lure pair ratings between context conditions, specifically between original and novel context. A follow up paired sample t-test revealed only two effects. Familiarity ratings for lures in original context (M = 2.95, SD = 1.33) and repaired context (M = 3.02, SD = 1.25) was greater than in novel context (M = 2.21, SD = 1.46), t (88) = 4.35, p <0.00 and t (88) = 5.21, p <0.00.

Overall, results suggest although participants have a higher subjective rating of familiarity for intact pictures compared to lures regardless of context and age, context condition does seem to matter for lure pictures. Specifically, participants have higher subjective ratings of familiarity for lures shown in the original context compared to the novel context. Repaired context and age have no significant difference in any condition for the lure picture pairs. This pattern of familiarity for intact vs. lure picture pairs is different from recollection for intact vs. lure picture pairs where we found no significant effects. Considering this pattern was not

replicated in the recollection item type, we think reinstating context to enhance memory illusions for lures may have had a bigger effect on familiarity which generally does not focus on details, thus easier to distort, compared to recollection.

Familiarity for Responses (Old vs. New) for Picture Pairs by Age. We wanted to investigate whether subjective ratings of the feeling of "oldness" differed in terms of age and context conditions for items participants deemed as "Old", in other words, they think they have seen before. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Response: Old vs New) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor was conducted to investigate the effect of context conditions, age and response type on familiarity ratings. This analysis considered whether the targets or lures were correctly endorsed as "Old" or "New"

Mauchly's test revealed assumption of sphericity has been violated for context, χ^2 (2) = 7.39, ε = 0.92 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.92>0.75) p = 0.03 and for context by item χ^2 (2) = 19.42, ε = 0.83 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.83>0.75), p <0.00, therefore degrees of freedom were corrected using Huyunh-Feldt estimates of sphericity.

Main effects: There was a main effect of context, F (1.91, 165.99) = 15.04, p <0.00 with a paired sample t-test showing responses in original context (M = 2.85, SD = 0.95) were greater compared to repaired context (M = 2.79, SD = 1.13), t (58) = 2.30, p <0.03 and responses in repaired context was greater compared to novel contexts (M = 2.26, SD = 1.16), t (58) = 3.70, p <0.00.

There was a main effect of response F (1, 87) = 231. 46, $p < 0.00$ with paired sample t-test showing old responses (M = 4.32, SD = 1.45) had higher ratings than new responses (M = 2.50, SD = 0.91), t (88) = 12.86, p <0.00.

Interaction: There was a significant 2-way interaction for response by age F (1, 87) = 11.137, p <0.00 and for context by response F (1.71, 148.92) = 15.56, p <0.00. The 2-way interactions were followed by a significant 3-way interaction of context by response by age, F (1.71, 148.92) = 3.63, p = 0.04 (Fig 5).

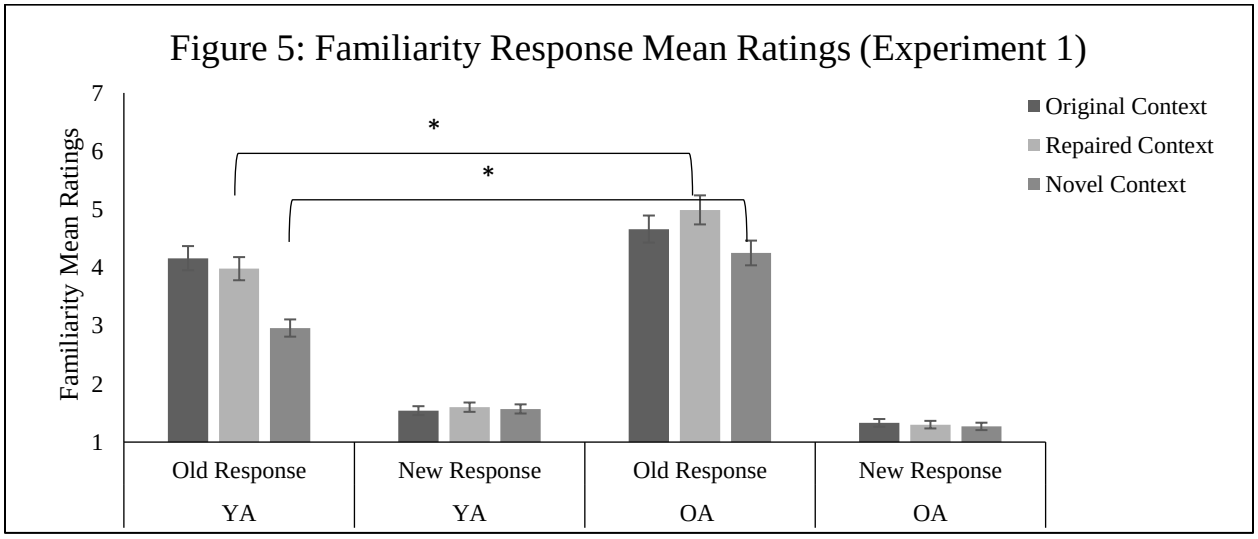


Figure 5: Mean familiarity ratings for pairs between old response and new response across context and age. Scale range from 1-7. There was a significant 3-way interaction of context by age by response, F (1.71, 148.92) = 3.63, p = 0.04

To understand this interaction, we first did two separate 3 (Context: Old, Repaired, Novel) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and age a between measure on Old items and on New items as the dependent variable for each test respectively to see whether ratings differ in old vs. new responses between younger and older adults.

For new responses, there were no significant effects. For old responses, Mauchly's test revealed the assumption of sphericity has been violated for context, χ^2 (2) = 25.56, ε = 0.80 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.80>0.75) *p* <0.00, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity.

For old responses, we found a 2-way interaction of context by age, *F* (1.63, 142.19) = 3.79, *p* <0.03. To understand this interaction, first, an independent t-test revealed two effects. Across old responses older adults had significantly (*M* = 4.66, *SD* = 1.56) higher responses compared to younger adults (*M* = 4.16, *SD* = 1.59) for pairs in repaired context, *t* (87) = 2.93, *p*<0.00. Older adults had significantly (*M* = 4.25, *SD* = 1.75) higher responses compared to younger adults (*M* = 2.96, *SD* = 1.44) novel context *t* (87) = 3.50, *p*<0.00. We then split the file into young and old age groups to further understand how old responses differ between younger vs. older adults. We re-ran the ANOVA testing for simple effects.

For younger adults, Mauchly's test revealed assumption of sphericity has been violated for context, χ^2 (2) = 0.89, ε = 0.80), *p* = 0.02, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. There was a main effect of context, *F* (1.69, 48.87) = 10.99, *p* <0.00. A follow up paired sample t-test revealed for younger adults, familiarity ratings for picture pairs in original context (*M* = 3.41, *SD* = 1.18) and repaired context (*M* = 3.21, *SD* = 1.31) was higher compared to those in novel context (*M* = 2.64, *SD* = 1.28), *t* (29) = 3.28, *p* <0.00 and *t* (29) = 2.70, *p* = 0.01.

In older adults, Mauchly's test revealed assumption of sphericity has been violated for context, χ^2 (2) = 18.36, ε = 0.78), *p* < 0.00, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. There was a main effect of context, *F* (1.60, 93.04) = 11.27,

p <0.00. A follow up paired sample t-test revealed for older adults, familiarity ratings for picture pairs in original context (M = 3.61, SD = 1.06) and repaired context (M = 3.73, SD = 0.94) was higher compared to those in novel context (M = 3.36, SD = 1.281.07), t (58) = 2.33, p <0.02, t (58) = 3.87, p < 0.00.

Covariates: The effect of age was still significant after controlling for covariates, F (1, 86) = 9.35, p<0.00.

Overall, results suggest interaction for familiarity ratings for picture pairs are mainly driven by response type and context type. Older adults give higher ratings for pairs in a repaired and novel context compared to younger adults. However, both younger and older adults have higher subjective ratings of familiarity for picture pairs they identify as "Old" if they were presented in an original and repaired context compared to the novel context in the cued recall phase.

Recollection for Context Only Condition by Age

As a reminder, all three context scene conditions from cued recall were considered as "Old" in the context only condition phase since participants have seen it before in the cued recall test. At this stage, they were intermixed with a new novel context condition (termed as "Novel 1" in the analysis) which was only reserved for old/new judgments phase at context conditions. This new context was deemed as "novel" since participants have never seen these contexts before at the cued-recall test. Since in this phase, we only showed the black and white context scenes without pairs, analysis can only be done for recollection/familiarity for response type (Old vs. New) from the Old/New Judgment condition in the context condition (read Procedure

section/Fig. 1 for details) and not for item type that requires pairs to be presented, which is not the case at this stage.

Recollection for Response (Old vs. New) for Context Condition by Age. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings of the number of details recollected differed in terms of age and context conditions for only context scenes. In other words, do participants pay a high level of attention to context (in turn have higher recollected details for "Old" scenes) and how does it differ for different age groups? To understand this question, we conducted a 4 (Context: Old vs. Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as "Old" or incorrectly endorsed as "New."

Main effects: There was a main effect of response, F (1, 87) = 72.56, p <0.00 and paired sample t-test revealed old responses (M = 3.17, SD = 1.51) have higher ratings than new (M = 1.64, SD = 1.13) responses, t (88) = 8.56, p <0.00.

Interaction: There was a significant 2-way interaction for context by response, F (3, 261) = 3.50, p = 0.02. This was followed by a significant 3-way interaction of context by response by age F (3, 261) = 5.46, p <0.00 (Fig 6).

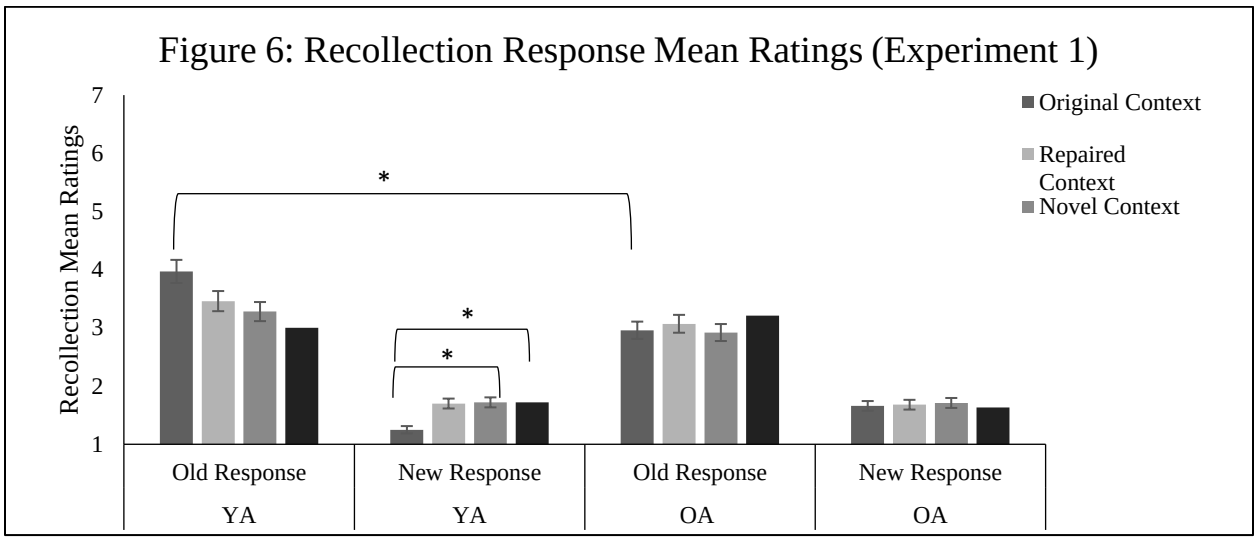


Figure 6: Mean recollection ratings for context between old response and new response across context and age. Scale range from 1-7. There was a significant 3-way interaction of context by age by response, F (3,261) = 5.46, p<0.00.

To understand the interaction further, we first did two separate 4 (Context: Old, Repaired, Novel, Novel 1) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and age a between measure on Old items and on New items as the dependent variable for each test respectively to see whether ratings differ in old vs. new responses.

For old items, we found a context by age interaction, F (3, 261) = 3.20, p = 0.02. To understand this interaction, first, an independent t-test revealed across old responses, younger adults had significantly higher responses compared to older adults for pairs in the original context, t (87) = 2.39, p = 0.02. We then split the file into young and old age groups to further understand whether old responses differ between younger vs. older adults. We re-ran the ANOVA testing for simple effects. Within old responses, there was a significant main effect of context in younger adults, F (3, 87) = 4.13, p <0.00 such that younger adults had higher recollection ratings for old responses for original context (M = 3.97, SD = 1.86) compared to

novel context (M = 3.27, SD = 2.10) at cued recall, t (29)=2.52, p = 0.02 and novel context in the context (M = 3.00, SD = =1.24) condition only, t (29) = 4.02, p <0.00.

Within new responses, Mauchly's test revealed assumption of sphericity has been violated, χ^2 = (5) = 24.13, p < 0.00, ε = 0.64 (<0.75) therefore degrees of freedom were corrected using Greenhouse Geisser estimates of sphericity. We found a significant interaction of context by age, F (2.46, 213.63) = 3.00, p = 0.04. Independent t-test revealed no age effects for scenes with new response across four types of context, p = 0.11, 0.94, 0.99, 0.75. We split the file into young and old age groups to see how the ratings differed individually between younger and older adults. Results revealed there was a marginal significant main effect of context in younger adults, F (1.92, 55.59) = 3.14, p = 0.05 such that younger adults had higher recollection ratings for new responses for novel context (M = 1.71, SD = 1.22) at cued recall compared to original context (M = 1.24, SD = 1.05), t (29) = 2.47, p = 0.02 and higher for novel context (M = 1.72, SD = 1.20) in the context condition only compared to original context (M = 1.24, SD = 1.05), t (29) = 2.43, p < 0.02.

For older adults, there were no significant effects of old responses, F (3, 174) = 0.55, p = 0.65 or for new responses, F (3, 174) = 0.22, p = 0.88 in recollection ratings for context only condition.

Covariates: The age effect was still significant after controlling for covariates, F (3, 258) = 5.09, p<0.00.

Overall results suggest younger adults had higher subjective memory for original scenes they responded as Old, compared to older adults. Scenes responded as New had no age effect. The three-way interaction was driven by younger adults' recollection ratings for context

conditions and response type, although in opposite patterns. Younger adults seem to have higher subjective recollection for original background scenes compared to novel background scenes, but only if they identified the context to be “Old”. When younger adults identify the background scenes as “new”, they seem to give higher recollection ratings for novel background scenes compared to an original background scene. The significant effect seems to hold even after controlling for covariates suggesting it is due to age differences. One reason could be younger adults' subjective ratings of recollection seem to enhance whenever they try to identify the scene as old or new. Perhaps trying to recall some part of scenery increases their subjective feelings of recollection, even if it is incorrect. The other reason could be younger adults simply did not calibrate their responses for old/new judgment with their subsequent recollection ratings. On the other hand, context or response type had no significant effect on older adults’ recollection ratings for context scenes. This may suggest older adults perhaps have less recollected details but more of the gist of the scenes. Perhaps they use more of familiarity mechanisms instead of recollection mechanisms to access their subjective feeling of any details.

Familiarity for Context Condition by Age

Familiarity for Response (Old vs. New) for Context Condition by Age. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings for a feeling of “oldness” for context scenes differed in terms of age and context conditions. In other words, do participants pay a high level of attention to context (in turn have a higher feeling of oldness for “Old” scenes) and how does it differ for different age groups? To understand this question, we conducted a 4 (Context: Old vs. Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Age: Young vs. Old)

mixed ANOVA with context and response as a within-subject and age as a between-subject factor. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as "Old" or incorrectly endorsed as "New."

Main Effects: There was a significant main effect of response F (1, 87) = 114.68, p <0.00 and a paired sample revealed old responses (M = 3.20, SD = 1.52) had higher ratings than new responses (M = 1.40, SD = 0.93), t (88) = 10.87, p <0.00.

Interaction: Results revealed a significant three-way interaction of context by response by age F (3, 261) = 4.94, p <0.00 (Fig 7).

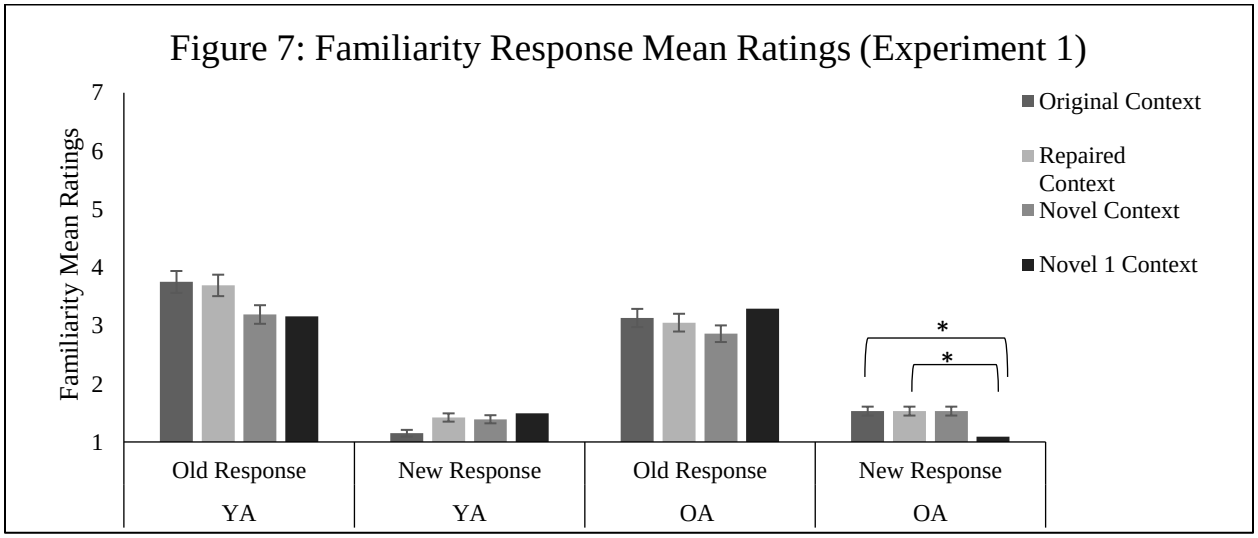


Figure 7: Mean familiarity ratings for context between old response and new response across context and age. Scale range from 1-7. There was a significant 3-way interaction of context by age by response, F (3, 261) = 4.94, p<0.00.

To understand the interaction further, we first did two separate 4 (Context: Old, Repaired, Novel, Novel 1) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and

age a between measure on Old items and on New items as the dependent variable for each test respectively to see whether ratings differ in old vs. new responses.

For old items, we found no significant effects. For new responses, Mauchly's test revealed assumption of sphericity has been violated, $\chi^2 = (5) = 29.73$, $p < 0.00$, $\varepsilon = 0.83$ (>0.75) therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. We found a context by age interaction, F (2.62, 227.93) = 7.95, p <0.00. Independent sample t-test revealed no age effects between new response types across each context. We then split the file into young and old age groups to further understand how new responses differ between younger vs. older adults. We re-ran the ANOVA testing for simple effects.

Within new responses, Mauchly's test revealed assumption of sphericity has been violated for older adults, Mauchly's test revealed assumption of sphericity has been violated, $\chi^2 = (5) = 34.56$, $p < 0.00$, $\varepsilon = 0.70$ (<0.75) therefore degrees of freedom were corrected using Greenhouse Geisser estimates of sphericity. There was a significant main effect of context in older adults, F (2.10, 121.77) = 9.87, p <0.00 such that older adults had higher familiarity ratings for new responses for original context (M = 1.54, SD = 1.11) compared to novel context (M = 1.09, SD = 1.13) at context condition only, t (58)=3.79, p <0.00 and higher for repaired context compared (M = 1.53, SD = 1.20) to novel context (M = 1.09, SD = 1.13) in the context condition only, t (58) = 3.60, p <0.00.

For younger adults, there were no significant effects for new responses, F (2.12, 61.59) = 2.68, p = 0.07 (Greenhouse Geisser estimates were used due to violation of sphericity χ^2 (5) = 17.37, p <0.00, $\varepsilon = 0.71$) in familiarity ratings for context condition.

Covariates: The effect of age was still significant after controlling for the covariates F (3, 258) = 4.86, p<0.00.

Overall results suggest the interaction for familiarity ratings in the context only condition was mainly driven by older adults' ratings for scenes they deemed as new or never seen before. This effect exists even after controlling for covariates suggesting the effect is not due to third variable factors. Unlike recollection, older adults seem to use some form of familiarity mechanism to access their subjective feeling of oldness for new scenes. Results may suggest that when older adults are explicitly saying "new" to a scene, this explicit identification in the next phase can implicitly bring some awareness of partial details that may seem to drive them to give higher ratings for original or repaired scene compared to a novel scene they have never seen before. In other words, as older adults have a lower familiarity for new responses compared to old responses (since old responses overall had higher familiarity compared to new responses regardless of age, t (88) = 10.87, p <0.00), it seems to suggest context types specifically come into play when older adults have low subjective feelings of familiarity. On the other hand, context or response type does not seem to use younger adults' familiarity ratings for background scenes, suggesting they may use more of a recollection mechanism to access their subjective feeling of oldness as shown in the previous section.

Exploratory Analysis of FOK measures with Recollection and Familiarity Ratings with Context Conditions

Since our main interest of measure was FOK with age and context, and we did not find any age or context effect, we wanted to understand further what drove the FOK ratings in our experiment if any. Specifically, since we found some age effects in recollection and familiarity

for context conditions, we wanted to examine whether these subjective aspects of memory drove FOK ratings. To test this question, we conducted a linear regression with FOK original, repaired, and novel context as DVs and Recollection and Familiarity for every three contexts as IV. Each IV was entered separately as an initial correlation between recollection and familiarity showed high significant collinearity among each other.

Recollection Original Context Predict FOK Original Context

A linear regression established that recollection ratings in original context could significantly predict FOK ratings in original context, $\beta = 0.39$, $SE = 1.77$, $t\ (87) = 3.96$, $p < 0.00$, $R^2 = 0.15$, $F\ (1, 87) = 15.68$, $p < 0.00$. Recollection ratings in original context accounted for 15.3% of the explained variability in FOK ratings in original context (Fig 8).

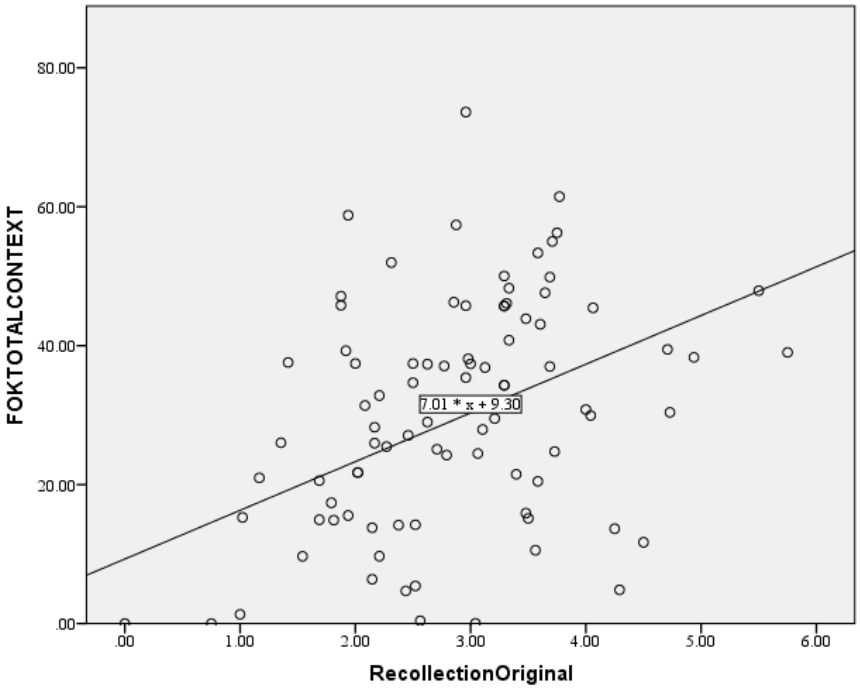


Figure 8: Scatterplot of recollection original context ratings predicting FOK ratings

Recollection Repaired Context Predict FOK Repaired Context

A linear regression established that recollection ratings in repaired context could statistically significantly predict FOK ratings in repaired context, $\beta = 0.31$, $SE = 1.79$, t (87) = 3.02, p <0.00, $R^2 = 1.00$, F (1, 87) = 9.13, p <0.00. Recollection ratings in repaired context accounted for 9.50 % of the explained variability in FOK ratings in repaired context (Fig 9).

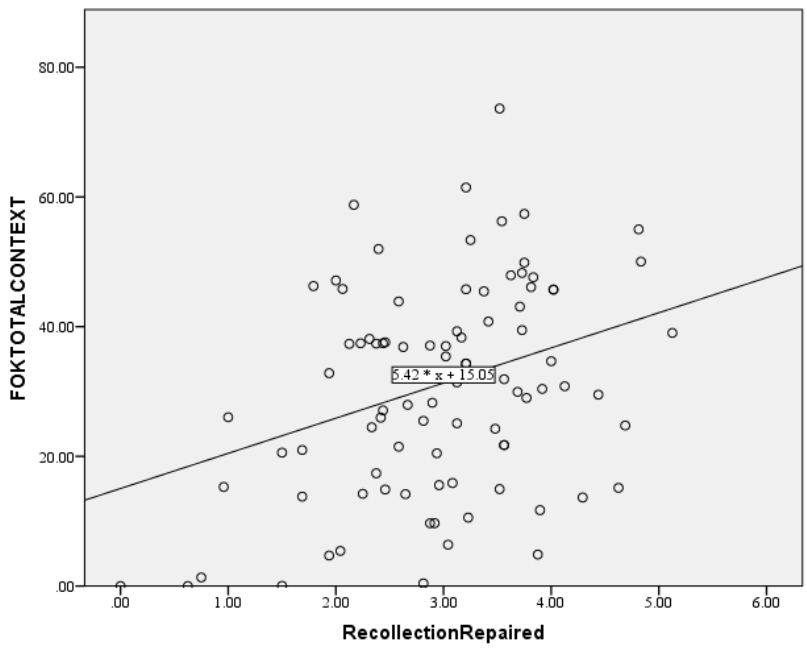


Figure 9: Scatterplot of recollection repaired context ratings predicting FOK ratings

Recollection Novel Context Predict FOK Novel Context

A linear regression established that recollection ratings in novel context could statistically significantly predict FOK ratings in novel context, $\beta = 0.29$, $SE = 1.83$, t (87) = 2.79, p <0.00, R^2 = 0.09, F (1, 87) = 7.80, p <0.00. Recollection ratings in novel context accounted for 8.20 % of the explained variability in FOK ratings in novel context (Fig 10).

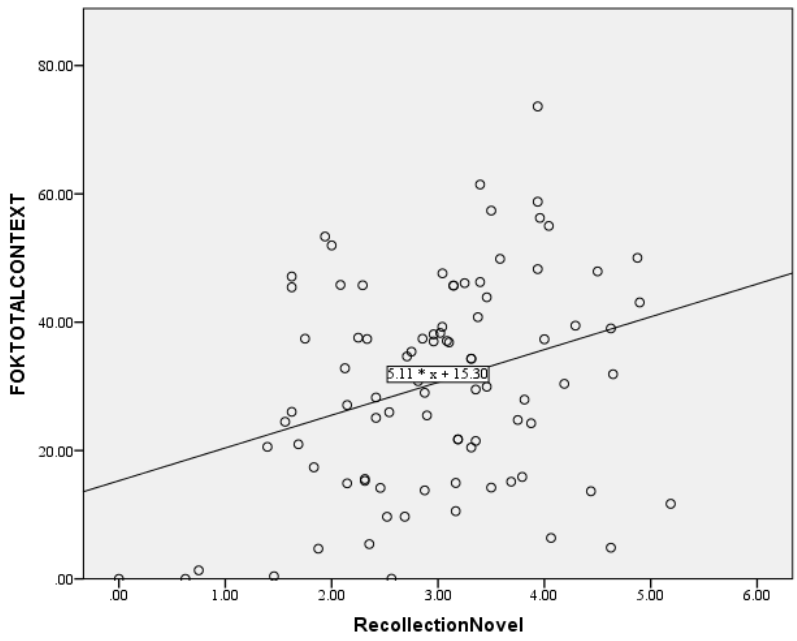


Figure 10: Scatterplot of recollection novel context ratings predicting FOK ratings

Familiarity Original Context Predict FOK Original Context

A linear regression established that Familiarity ratings in original context could statistically significantly predict FOK ratings in original context, $\beta = 0.47$, $SE = 1.59$, $t\ (87) = 4.99$, $p < 0.00$, $R^2 = 0.22$, $F\ (1, 87) = 24.90$, $p < 0.00$. Familiarity ratings in original context accounted for 22.30% of the explained variability in FOK ratings in original context (Fig 11).

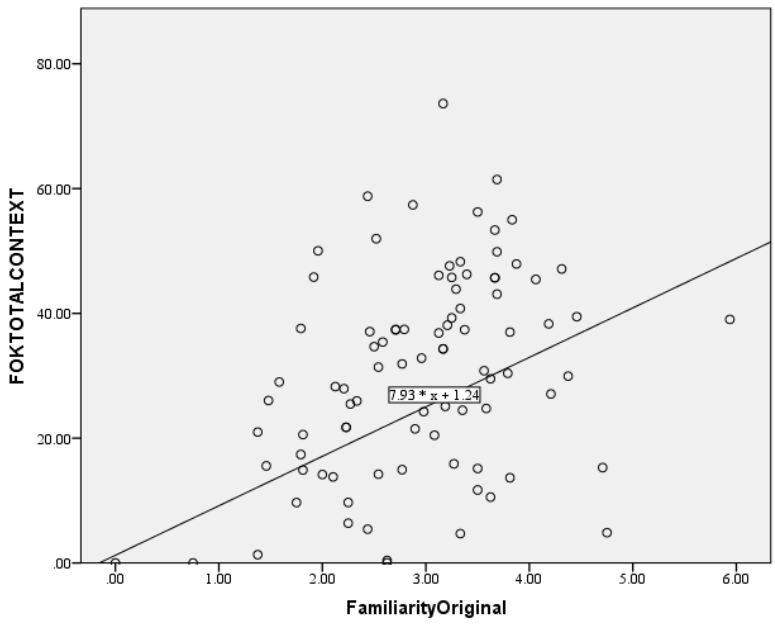


Figure 11: Scatterplot of familiarity original context ratings predicting FOK ratings

Familiarity Repaired Context Predict FOK Repaired Context

A linear regression established that Familiarity ratings in Repaired context could statistically significantly predict FOK ratings in Repaired context, $\beta = 0.47$, $SE = 1.48$, $t\ (87) = 5.00$, $p < 0.00$, $R^2 = 0.22$, $F\ (1, 87) = 25.05$, $p < 0.00$. Familiarity ratings in Repaired context accounted for 22.40% of the explained variability in FOK ratings in Repaired context (Fig 12).

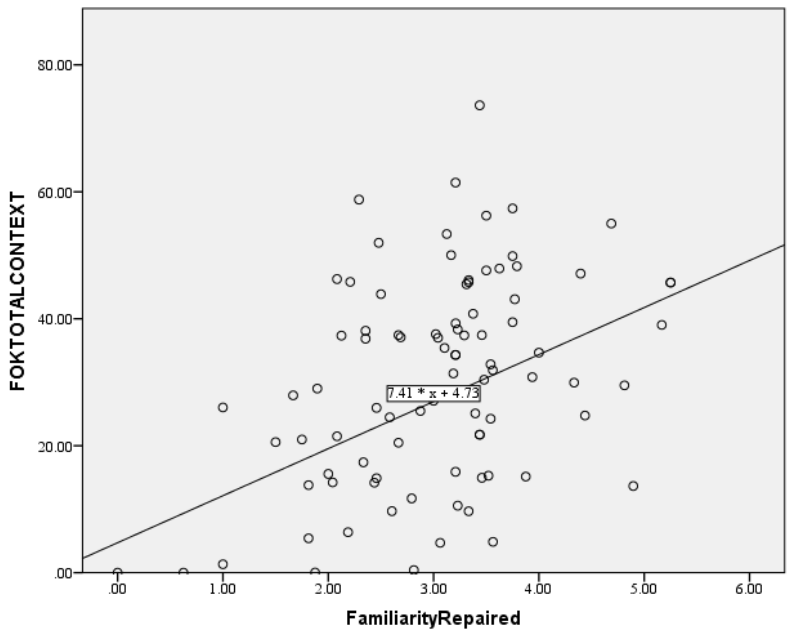


Figure 12: Scatterplot of familiarity repaired context ratings predicting FOK ratings

Familiarity Novel Context Predict FOK Novel Context

A linear regression established that Familiarity ratings in Novel context could statistically significantly predict FOK ratings in Novel context, $\beta = 0.36$, $SE = 1.50$, $t\ (87) = 3.55$, $p < 0.00$, $R^2 = 0.13$, $F\ (1, 87) = 12.59$, $p < 0.00$. Familiarity ratings in Novel context accounted for 12.60% of the explained variability in FOK ratings in Novel context (Fig 13).

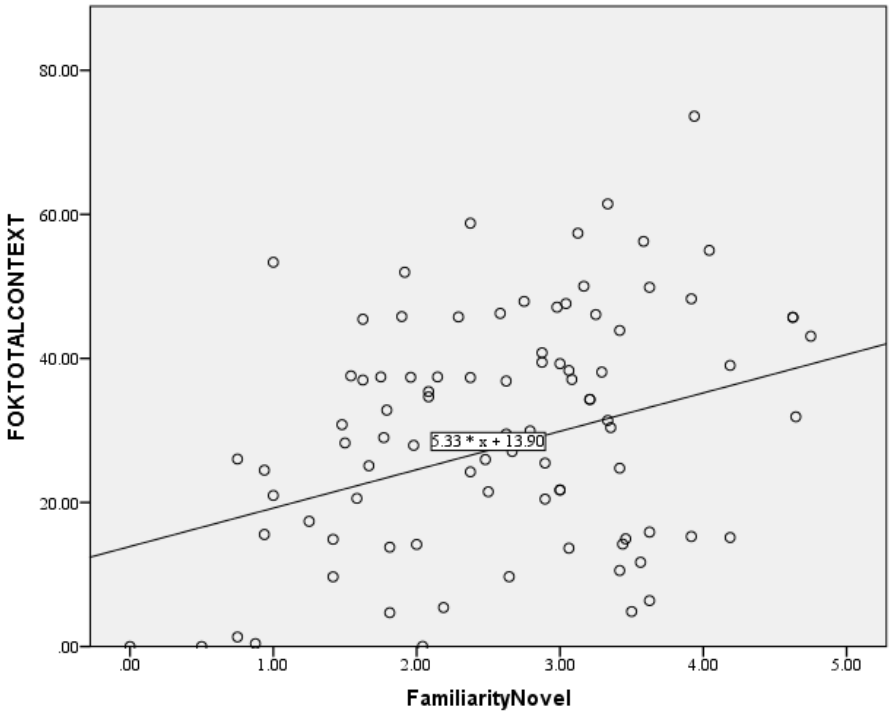


Figure 13: Scatterplot of familiarity novel context ratings predicting FOK ratings

Results seem to suggest both recollection and familiarity ratings predict FOK ratings for each type of context. Recollection original context ratings seem to predict a higher amount of explained variability for FOK original context ratings than other conditions. Familiarity original and repaired context ratings seem to predict a higher amount of explained variability for predicting FOK original and repaired context, compared to novel context pairing. Looking at overall percentages for subjective ratings, results show familiarity overall (across context) predicts a higher percentage of FOK rating variability compared to recollection ratings overall (across context). We suggest that perhaps older adults' reliance more on familiarity with age compared to younger adults may seem to have accounted for this effect.

Interaction of Age by Familiarity

To test the large effect of age groups on overall familiarity as a possible explanation on overall FOK ratings as mentioned in the previous section, we re-did the analysis with additional two predictors, age groups, and age by familiarity interaction for each context condition. Only significant effects are mentioned. We found for original context, only the age by familiarity interaction was significant, $\beta = 2.09$, $SE = 5.81$, $t\ (85) = 5.36$, $p < 0.00$, $R^2 = 0.37$, $F\ (3, 85) = 16.26$, $p < 0.00$ (Fig 14). Familiarity with Age interaction in the original context accounted for 37% of the explained variability in FOK ratings.

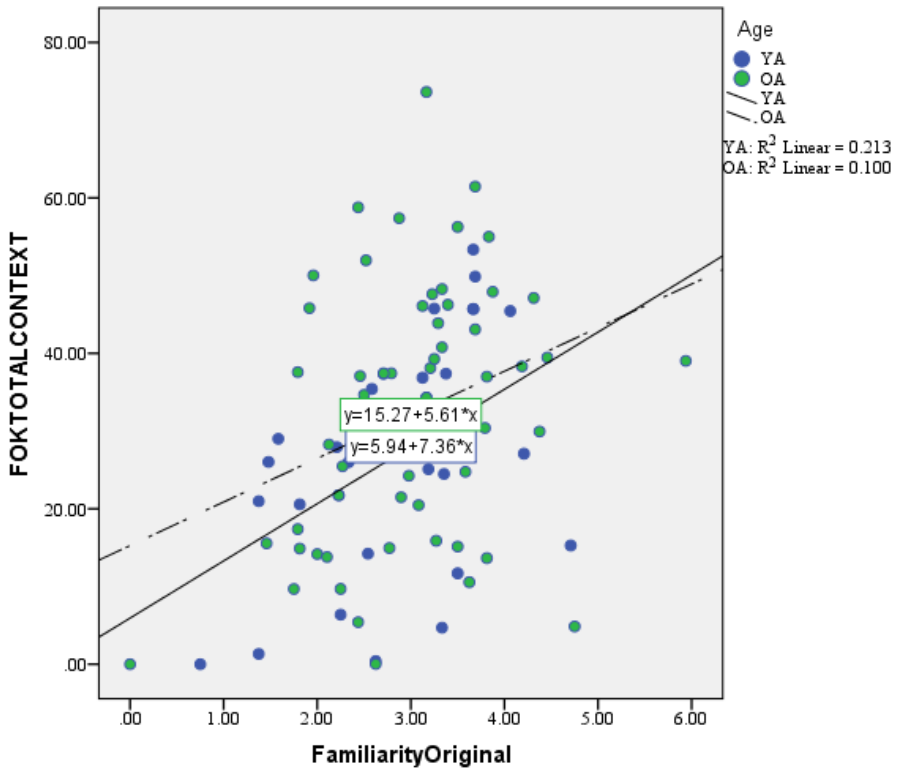


Figure 14: Scatterplot of familiarity by age interaction in original context predicting FOK ratings

For repaired context, only the age by familiarity interaction was significant, $\beta = 0.67$, $SE = 3.31$, t (85) = 3.00, p <0.00, $R^2 = 0.26$, F (3, 85) = 9.68, p <0.00 (Fig 15). Familiarity by Age interaction in repaired context accounted for 26% of the explained variability in FOK ratings.

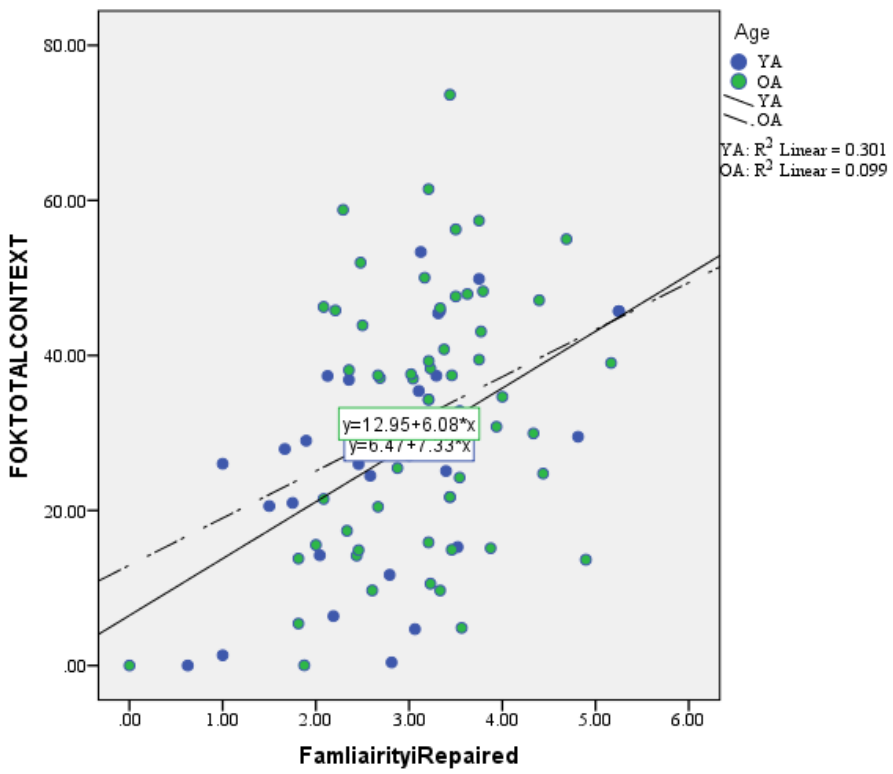


Figure 15: Scatterplot of familiarity by age interaction in repaired context predicting FOK ratings

For novel context, only the age by familiarity interaction was significant, $\beta = 1.04$, $SE = 6.02$, t (85) = 2.38, $p = 0.02$, $R^2 = 0.18$, F (3, 85) = 6.31, p <0.00 (Fig 16). Familiarity by Age interaction in novel context accounted for 18% of the explained variability in FOK ratings.

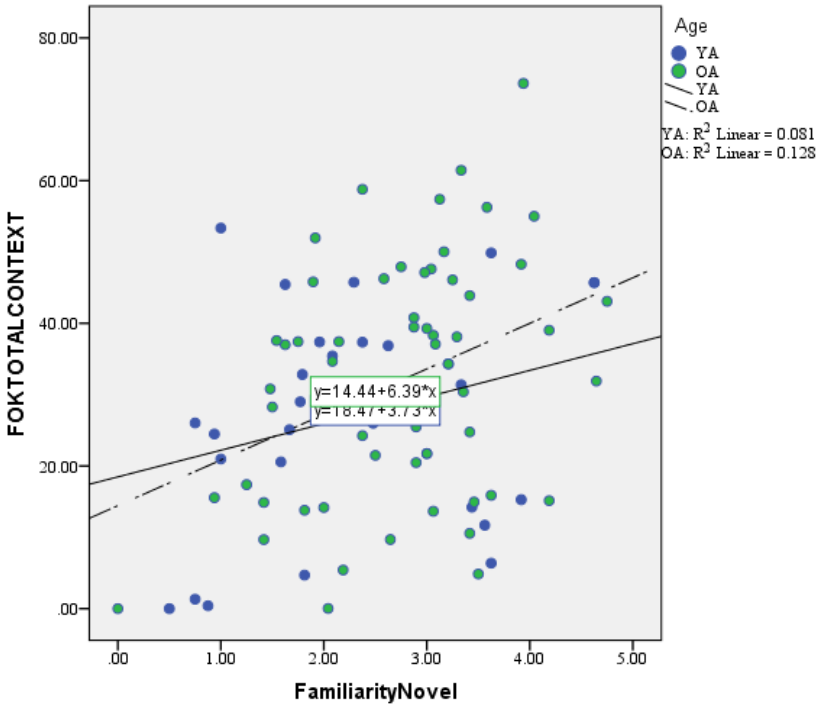


Figure 16: Scatterplot of familiarity by age interaction in the novel context predicting FOK ratings

Overall results suggest interaction of familiarity rating with age predicted overall FOK ratings. On one hand, a significant interaction means a better fit to our data. We can see that from R^2 value increase with and without the interaction (from ~22% to an increase of 26-37% for original and repaired context and from ~13% to 18% for novel context). On the other hand, a significant interaction also means that we have uncertainty about the main effect predictions, i.e. familiarity ratings, on FOK ratings. In sum, while we cannot generalize whether it is only familiarity ratings by itself that can predict FOK ratings, we can imply that age in addition to familiarity also has a role to play in driving FOK ratings.

Analysis by Recruitment for Older Adults (In-person versus M Turk)

Demographics and Covariate Analysis for Recruitment

See Table 2 for demographics. An independent sample t-test was conducted to investigate whether M Turk older adult samples differed from the in-person older adult sample in terms of demographics. The tt-test revealed in-person sample had higher number of years in education level, t (57) = 2.14, p = 0.04 and had lower verbal scores t (57) = 1.96, p = 0.05 compared to M Turk sample. From correlation, we found the two scores were not significantly correlated with each other, so a composite score was not created, unlike the whole sample. Instead, both covariates were entered together in the analysis, and a similar analysis as before was conducted whenever there was a significant recruitment effect in the main analysis.

A similar analysis for FOK, Recall, Discrimination scores, Recollection-Familiarity measures was conducted as above replacing Age group with Recruitment group for older adults (In-person vs. M Turk) to investigate whether recruitment differences for older adults influenced experimental measures.

Feeling of Knowing (FOK) Mean Ratings by Recruitment

We wanted to investigate whether metamemory, aka FOK mean ratings differed in terms of recruitment and context conditions. To understand this question, we conducted a 2 (Response: Recall vs. Unrecall) x 2 (Recruitment: OA M Turk vs. OA in-person) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with response and context as within-subject and Recruitment as a between-subject factor.

Main Effect: There was a main effect of response type F (1, 57) = 90.94, p<0.00 where recalled items (M = 49.07, SD = 28.11) had higher FOK ratings than unrecalled items (M = 15.08, SD = 14.59), t (58) = 8.79, p<0.00.

Interaction: Results revealed a significant interaction of response type by Recruitment, F (1, 114) =10.51, p< 0.00 (Fig 17).

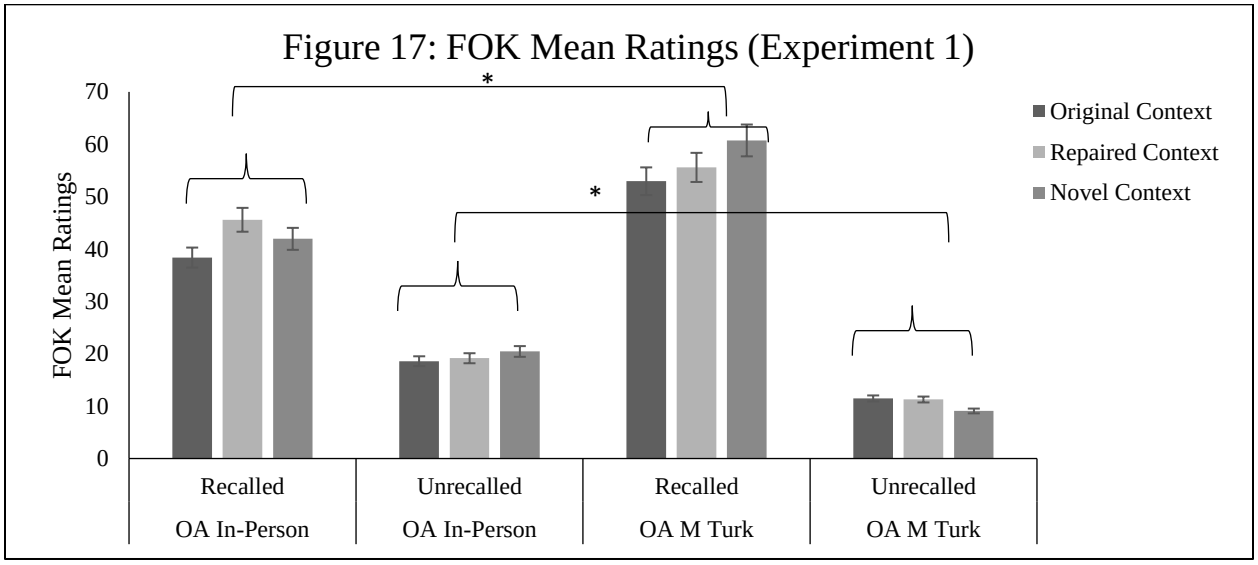


Figure 17: Mean Feeling of Knowing (FOK) ratings between recalled and unrecalled items across context and recruitment. Scale range from 0-100. There was a 2-way interaction of response by recruitment, F (1,114) = 10.51, p< 0.00.

An independent sample t-test with recruitment as a group variable and FOK response type as dependent variable revealed for recalled responses only, M Turk older adults had higher FOK ratings (M = 56.42, SD = 25.18), compared to in-person older adults (M = 41.96, SD = 29.36), t (57) = 2.03, p = 0.05. On the other hand, in-person older adults had higher FOK ratings for unrecalled items (M = 19.40, SD = 17.88) compared to M Turk older adults (M = 10.61, SD = 8.33), t (57) = 2.41, p = 0.02.

Covariates: After controlling for the number of years of education and verbal score, the recruitment effect was still significant F (1,55) = 12.12, p<0.00.

Overall results suggest although there is no context effect on FOK ratings, response type differed depending on whether older adults were from the M Turk sample or in-person, even after controlling for differences in education or verbal scores. M Turk sample was overconfident for recalled items compared to in-person sample but underconfident for unrecalled items. This may also suggest that although the absence of context influence on FOK ratings in the whole sample and within older adults may still be generalizable, the absence of response effects in the whole sample may not be generalizable to the population due to recruitment differences of ratings within older adults.

Cued Recall by Recruitment

We wanted to investigate whether objective memory such as cued recall would differ with recruitment and context conditions. To understand this question, we conducted a 2 (Recruitment: OA M Turk vs. OA in-person) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with Recruitment as a between-subject factor and context as a within-subject factor.

Main effects: No significant effects of context were found, F (2, 114) = 0.02, p = 0.98 (Table 8). There was a significant effect of age such that M Turk sample had higher cued recall than OA-in person, F (1, 57) = 24. 55, p <0.00.

Table 8

Descriptive Statistics of Cued Recall Experiment 1 by Recruitment

	Recruitment	Mean	Std. Deviation	N
Recall_original	MturkOA	.45	.25	29.00
	In-person OA	.19	.19	30.00
	Total	.32	.26	59.00
Recall_repaired	MturkOA	.45	.29	29.00
	In-person OA	.20	.17	30.00
	Total	.32	.26	59.00
Recall_novel	MturkOA	.47	.25	29.00
	In-person OA	.17	.16	30.00
	Total	.31	.26	59.00

Interaction: No significant effects were found, context by recruitment, F (2, 114) = 0.80, p = 0.45.

Overall results suggest the type of context does not influence cued recall for older adults although in general, M Turk has higher cued recall than older adults in person. This is similar for the whole sample when older adults sample were combined and compared with younger adults. This may imply our results are generalizable, at least concerning this experiment as there is no recruitment difference.

Old/New Judgment by Recruitment

We also measured another kind of objective memory besides cued recall—participant's ability to discriminate between target pairs from lure pairs that were perceptually similar to target pairs. The purpose was to explore whether older adults' recruitment differences (in-person and M Turk sample) played a role in recognizing target pairs better compared to lure pairs across context conditions in the cued recall phase.

Thus, two measures were obtained from old/new judgment: hit and false alarms. In the picture pair phase, hit rates were calculated as the proportion of correctly selecting “old” for the original cue-original target pair. In context only condition phase, hit rates were calculated as the proportion of correctly selecting “old” for original, repaired, and novel contexts which they have seen at cued recall. In the picture pair phase, false alarms were calculated as the proportion of incorrectly selecting “old” for the cue-lure target pair. In context only condition phase, a false alarm was calculated as the proportion of incorrectly selecting “old” for the new novel context exclusive at the context only condition, which participants have never seen before at cued recall.

After each measure was collected, for the picture pair phase, we calculation a discrimination score (Hits-FA) for each context condition—original, repaired, and novel contexts, reported below. We did not calculate discrimination score for context condition only but reported hits and false alarms separately since hits and false alarms did not have the same context conditions due to presentation of context scenes without picture pairs at this stage.

Discrimination Score (Hits-False Alarms) for Picture Pairs by Recruitment. A 2 (Age: Young vs. Old) x 2 Context at Cued-Recall Test (Original, Repaired, Novel) mixed ANOVA with recruitment as a between-subject factor and context as a within-subject factor was conducted to test the effect of context and recruitment on participant’s discrimination ability between target vs. lure pairs.

Main effects: No significant effects were found, F (2, 114) = 0.61, p = 0.55 (Table 9).

Table 9

Descriptive Statistics for Discrimination Scores Experiment 1

	Recruitment	Mean	Std. Deviation	N
Pairs_Original	MturkOA	.31	.51	29.00
	In-person OA	.39	.28	30.00
	Total	.35	.41	59.00
Pairs_Repaired	MturkOA	.27	.28	29.00
	In-person OA	.32	.23	30.00
	Total	.302	.25	59.00
Pairs_Novel	MturkOA	.31	.34	29.00
	In-person OA	.38	.24	30.00
	Total	.34	.30	59.00

Interaction: No significant effects were found, context by recruitment, F (2, 114) = 0.04, p = 0.96.

Overall results suggest background type or recruitment does not influence memory discrimination for picture pairs or when older participants make errors in their memory discrimination, regardless of the sample type. This is similar for the whole sample when older adults sample were combined and compared with younger adults. This may mean our results are generalizable, at least concerning this experiment as there is no recruitment difference.

Hits for Context Condition Only by Recruitment. A 2 (Recruitment: OA M Turk vs. OA in-person) x 2 Context (Original vs. Novel) mixed ANOVA with Recruitment as a between-subject factor and context as a within-subject factor was conducted to test the effect of context and age on hit rates of contexts.

Main effects: No significant effects were found, F (2, 114) = 1.25, p = 0.29 (Table 10).

Table 10

Descriptive Statistics for Hits Context Condition Experiment 1 by Recruitment

	Recruitment	Mean	Std. Deviation	N
Context_Hits_original	MturkOA	.36	.30	29.00
	In-person OA	.43	.31	30.00
	Total	.40	.30	59.00
Context_Hits_repaired	MturkOA	.37	.28	29.00
	In-person OA	.39	.27	30.00
	Total	.38	.27	59.00
Context_Hits_novel	MturkOA	.33	.31	29.00
	In-person OA	.35	.27	30.00
	Total	.34	.29	59.00

Interaction: No significant effects were found, context by recruitment, F (2, 114) = 0.47, p = 0.63.

FA for Context Condition Only by Recruitment (Experiment 1). An independent t-test was conducted with Recruitment as a between-group as independent variable and ratings of false alarms for original context only as a dependent variable since false alarms cannot be computed in a novel context.

Main effects: No significant effects were found, t (57) = 0.09, p = 0.93 (Table 11).

Table 11

Group Statistics for FA Context Condition Experiment 1 by Recruitment

	Recruitment	N	Mean	Std. Deviation	Std. Error Mean
Context_FA_compt	In-person OA	30.00	.33	.20	.03
	MturkOA	29.00	.33	.24	.04

Interaction: Not applicable.

Overall results suggest within older adults, there is no effect of context or sample difference on discriminating between old or new scenes. Although when we combine older adult sample and compare with younger adults, we do see the main effect of context in the whole sample, the absence of such effect here may imply that the context effect may have been driven by the larger age difference between younger and older adults. Additionally, the interaction effects shown in the whole sample were driven mainly by younger adults' higher hit ratings.

Recollection for Picture Pairs by Recruitment

Recollection for Items (Target vs. Lures) for Picture Pairs by Recruitment. We wanted to investigate whether subjective ratings of the number of details recollected differed in terms of recruitment and context conditions for target pairs. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Recruitment: OA M Turk vs. OA in-person) mixed ANOVA with context and item as a within-subject and Recruitment as a between-subject factor to investigate the effect of context conditions and recruitment on recollection ratings for the pairs. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether recollection differs depending on whether it was a target or a lure, in addition to context and age conditions.

Main effect: Results revealed a main effect of context condition only, F (2, 114) = 194.16, $p < 0.00$ and a paired sample t-test revealed recollection ratings for items overall were greater for original context (M = 3.20, SD = 0.91) compared to repaired context (M = 1.32, SD = 0.11), t (58) = 13.36, $p < 0.00$ and novel context (M = 1.13, SD = 0.05), t (58) = 19.48, $p < 0.00$.

Rating in repaired ratings were marginally significantly greater compared to novel context, t (58) = 1.96, p =0.05.

Interaction: No significant effects were found, context by recruitment, F (2, 114) = 0.28, p = 0.76, item type by recruitment, F (1, 57) = 0.16, p = 0.69, context by item type, F (2, 114) = 1.71, p = 0.19, context by response by recruitment, F (2, 114) = 0.49, p = 0.61.

Overall results suggest regardless of recruitment and item type, older adults, in general, give higher subjective ratings of recollection for picture pairs.

Recollection for Response (Old vs. New) for Picture Pairs by Recruitment. We also wanted to investigate whether subjective ratings of the number of details recollected differed in terms of recruitment and context conditions for items participants deemed as "Old", in other words, pairs they think they have seen before. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Response: Old vs. New) X 2 (Recruitment: OA M Turk vs. OA in-person) mixed ANOVA with context and response as a within-subject and Recruitment as a between-subject factor was conducted to investigate the effect of context conditions and recruitment on recollected ratings for the pair responses given during their old/new judgment phase. This analysis considered whether the targets or lures were correctly endorsed as "Old" or "New".

Main Effect: There was a main effect of context F (2, 114) = 52.54, p< 0.00 and a paired sample t-test revealed recollection ratings for responses were only greater in original context (M = 3.97, SD = 1.28) compared to novel context (M = 3.08, SD = 0.90), t (58) = 8.42, p<0.00, and greater for repaired context (M = 4.04, SD = 1.10) compared to novel context, t (58)= 9.66, p<0.00.

Results also revealed a main effect of response, F (1, 114) = 212.27, p< 0.00. A paired sample t-test revealed overall old responses had higher ratings (M = 4.78, SD = 1.36) compared new responses (M = 1.45, SD = 0.84) t (58) = 17.60, p< 0.00.

Interaction: There was a significant 2-way interaction of response by context, F (1, 57) = 53.85, p< 0. 00. To understand this interaction, we did two separate mixed ANOVAs for old responses and new responses (Fig 18).

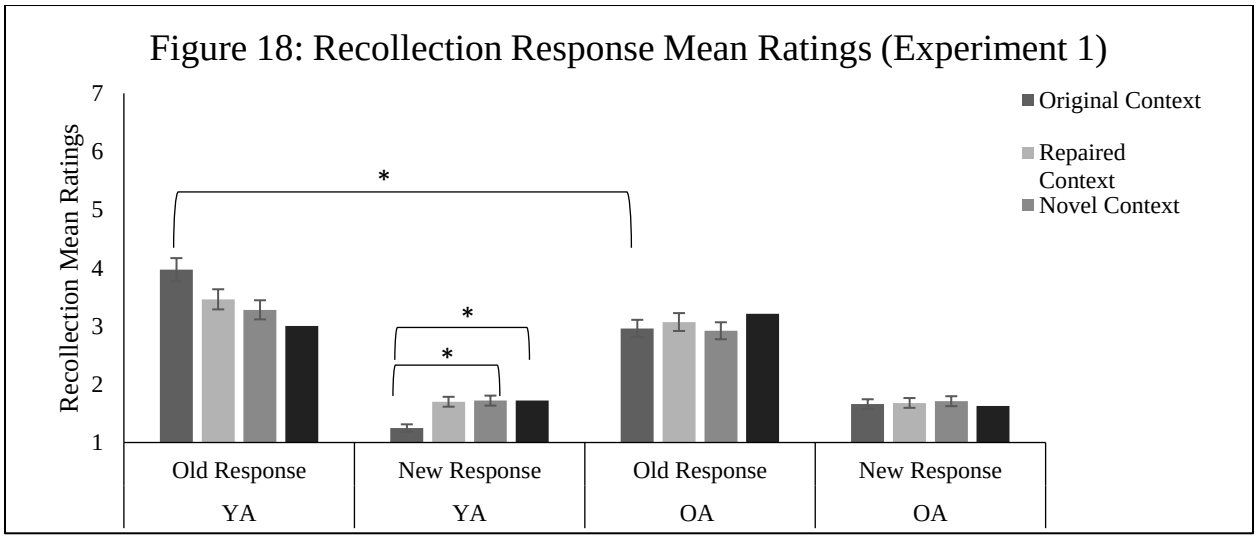


Figure 18: Mean recollection ratings for pairs between old response and new response across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of context by response for new responses, F (2,114) = 53.85, p<0.00.

We found for old responses there were no significant effects. For new responses we found a main effect of context, F (2, 114) = 102.90, p<0.00. A paired sample t-test revealed two effects only. For new responses, recollection ratings for were higher for original context (M = 4.89, SD = 1.44) compared to novel context (M = 3.00, SD = 1.05), t (58) =10.94, p<0.00. Recollection ratings for new responses were also higher for repaired context (M = 4.99, SD = 1.46) compared to recollection ratings for novel context, t (58) = 11.95, p< 0.00.

Overall, results show older adults, in general, have higher ratings for old scenes compared with new scenes for pairs that were marked as new in the old/new judgment phase.

Familiarity for Picture Pairs by Recruitment

Familiarity for Items (Target vs. Lures) for Picture Pairs by Recruitment. We wanted to investigate whether subjective ratings of the feeling of "oldness" differed in terms of recruitment and context conditions for target pairs. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Recruitment: OA M Turk vs. OA in-person) mixed ANOVA with context and item as a within-subject and Recruitment as a between-subject factor to investigate the effect of context conditions, Recruitment and item type on familiarity ratings. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether familiarity differs depending on whether it was a target or a lure, in addition to context and recruitment conditions.

Main effects: There was a main effect of context, F (2, 114) = 7.64, $p < 0.00$. A paired sample t-test revealed ratings for items in original context (M = 3.00, SD = 0.94) was greater for items in novel context, (M = 2.76, SD = 0.93), t (58) = 2.30, $p = 0.03$ and greater for items in repaired context (M = 3.15, SD = 0.87) compared to items in novel context, t (58) = 3.71, $p < 0.00$.

There was a main effect of item type F (1, 57) = 4.77, $p = 0.03$ with intact items having significantly higher familiarity ratings (M = 4.32, SD = 1.00) compared to lure items (M = 2.82, SD = 1.09), t (58) = 11.78, $p < 0.00$.

Interaction: Results revealed a significant interaction of context by item F (2,114) = 7.37, p = 0.01. To understand this interaction, we did two separate mixed ANOVAs with for intact and lure items (Fig 19).

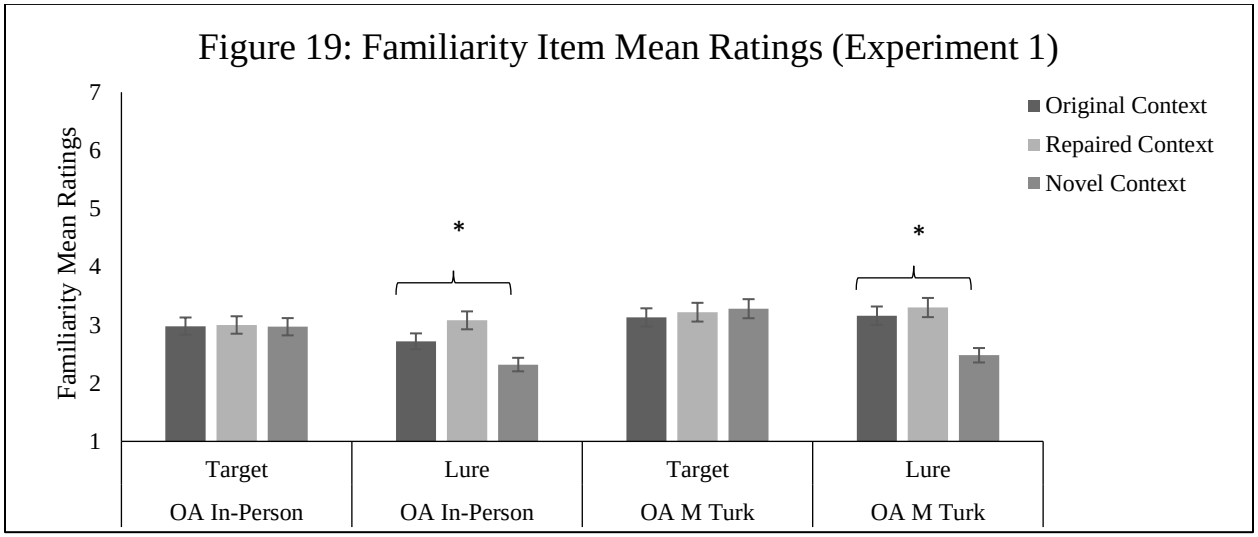


Figure 19: Mean familiarity ratings for pairs between target and lures across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of context by item for lures, F (2, 114) = 7.37, p < 0.00

We found for intact items there was no significant effects. For lure items we found a main effect of context, F (2, 116) = 9.70, p<0.00. A paired sample t-test revealed two significant effects. For lure items overall, familiarity ratings were higher for pairs in original context (M = 3.96, SD = 2.31) compared to pairs in novel context (M = 2.97, SD = 2.64), t (58) =2.83, p<0.00. Ratings were also higher for pairs in repaired context (M = 4.42, SD = 1.87) compared to pairs in novel context, t (58) = 4.40, p< 0.00.

Overall results suggest although recruitment has no influence on item type for familiarity ratings, context still influences older adults' ratings for item type, specifically for lures. Overall, older adults, in general, seem to mistakenly have higher familiarity for perceptually similar lures

in context scenes they have seen before such as original and repaired context compared to a novel context. Perhaps, regardless of explicit association of item-context, older adults may still take in some partial context cues to form an implicit association between the pairs and scenes. Thus, for a context they have seen before, lures may appear more familiar due to the previously made association with the target. Also, memory illusion in older adults for the high familiarity of lures may occur due to context reinstatement effects.

Familiarity Response (Old vs. New) for Picture Pairs by Recruitment. We wanted to investigate whether subjective ratings of the feeling of "oldness" differed in terms of recruitment and context conditions for items participants deemed as "Old", in other words, they think they have seen before. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor to investigate the effect of context conditions, age and response type on familiarity ratings. This analysis considered whether the targets or lures were correctly endorsed as "Old" or "New".

Main effect: There was a main effect of context, F (2, 114) = 67.60, $p < 0.00$. A paired sample t-test revealed ratings for overall responses ratings were greater for pairs in novel context, (M = 4.36, SD = 1.49) compared to pairs in original context (M = 3.03, SD = 0.96), t (58) = 10.06, $p < 0.00$, and pairs in repaired context (M = 3.20, SD = 0.87), t (58) = 8.01, $p < 0.00$. Pairs in repaired context was also marginally significantly greater compared to pairs in original context, t (58) = 2.03, $p = 0.05$.

There was also a main effect of response F (1,57) = 221.44, p< 0.00. A paired sample t-test revealed ratings for pairs with old responses (M = 4.63, SD = 1.42) were higher compared to pairs with new responses (M = 2.51, SD = 0.72), t (58) = 12.88, p< 0.00.

Interaction: Results revealed a significant interaction of context by response F (2,114) = 124.64, p< 0.00. To understand this interaction, we did two separate mixed ANOVAs for old responses and new responses (Fig 20).

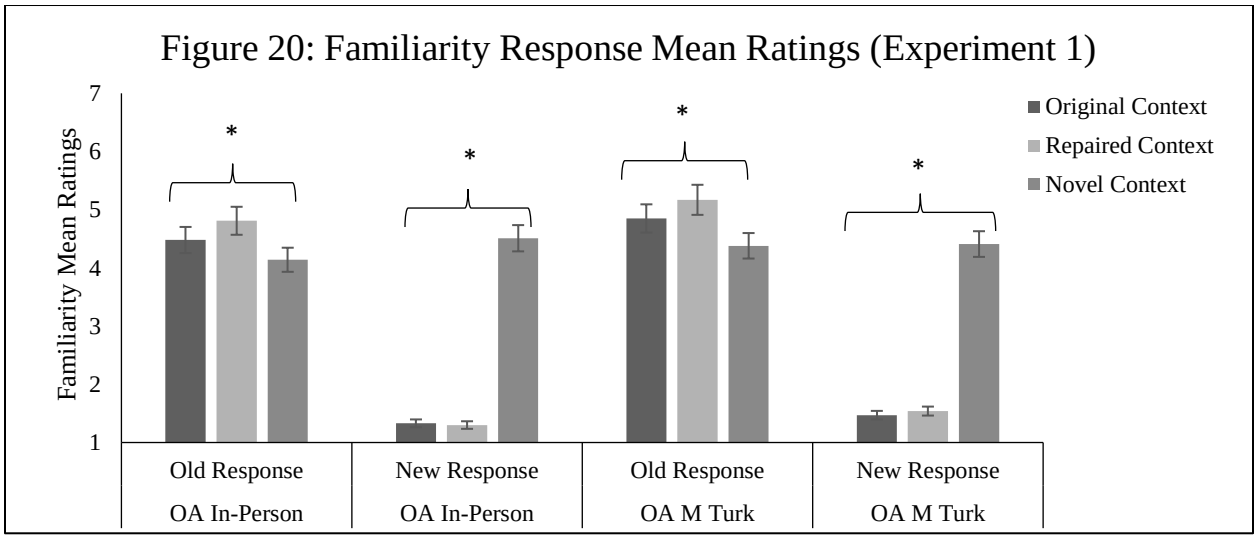


Figure 20: Mean familiarity ratings for pairs between old and new responses across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of context by response, F (2, 114) = 124.64, p < 0.00

For old responses we found a main effect of context, F (2, 116) = 11.27, p<0.00. A paired sample t-test revealed for old responses, familiarity ratings were higher for pairs in repaired context (M = 4.99, SD = 1.41) compared to pairs in original context (M = 4.65, SD = 1.56), t (58) =3.09, p<0.00 and in novel context (M = 4.25, SD = 1.75), t (58) =4.20, p<0.00. On the other hand, ratings for pairs in original context (M = 4.99, SD = 1.46) was only higher when compared to pairs in novel context, t (58) = 2.33, p< 0.02.

For new responses, we found a main effect of context, F (2, 11) = 152.27, p<0.00. A paired sample t-test revealed for new responses, familiarity ratings for pairs were higher in novel context (M = 4.46, SD = 1.55) compared to pairs in original context (M = 1.40, SD = 1.09), t (58) = 12.75, p<0.00 and in repaired context (M = 1.42, SD = 0.87), t (58) = 14.29, p<0.00.

Overall results show an interesting pattern where familiarity response ratings for old responses seem to higher for pairs in repaired>original>novel context. For new responses, familiarity response ratings seem to higher for pairs novel>original>repaired context scenes. It seems the interaction is mainly driven by ratings of pairs in repaired and novel contexts. For old responses, older adults seem more familiar to old items in a repaired context compared to a novel context while the opposite is true for new responses. One of the reasons for this pattern could be the case that older adults did not align their old/new judgment with their subjective ratings regardless of contextual scenes. Thus, even for responses, they said new which were in a novel context, they gave high familiarity ratings. In turn, context effects could have arisen due to chance errors. However, considering overall results where we saw no context effects for discrimination scores or cued recall, results here may suggest older adults could have been implicitly paid some attention to context scenes at the study phase to influence their subjective ratings of familiarity.

Recollection for Context Only Condition by Recruitment

As mentioned before, all three context scene conditions from cued recall were considered as “Old” in the context only condition phase since participants have seen it before in the cued recall test. At this stage, they were intermixed with a new novel context condition (termed as “Novel 1” in the analysis) which was only reserved for old/new judgments phase at context

conditions only. This new context was deemed as “novel” since participants have never seen these contexts before at cued-recall test. Since in this phase, we only showed the black and white context scenes without pairs, analysis can only be done for recollection/familiarity for response type (Old vs. New) from the Old/New Judgment condition in the context only condition (read Procedure section/Fig. 1 for details) and not for item type that requires pairs to be presented, which is not the case at this stage

Recollection for Response (Old vs. New) for Context Condition by Recruitment. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings of amount of details recollected differed in terms of recruitment and context conditions for only context scenes. In other words, does participants pay high level of attention to context (in turn have higher recollected details for “Old” scenes) and how does it differ for in-person and M Turk sample of older adults? To understand this question, we conducted a 4 (Context: Old, Repaired, Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Recruitment: In-person vs. M Turk) mixed ANOVA with context and response as a within-subject and Recruitment as a between-subject factor to investigate the effect of context conditions, recruitment and response type during old/new judgment phase on recollection ratings. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as “Old” or incorrectly endorsed as “New.”

Main effects: There was a main effect of response, F (1, 57) = 38.16, p <0.00 and paired sample t-test revealed old responses (M = 3.04, SD = 1.47) have higher ratings than new (M = 1.67, SD = 1.19) responses, t (58) = 6.04, p <0.00.

Interaction: There was a significant 2-way interaction for response by recruitment, F (1, 57) = 4.36, p = 0.04. An independent t-test revealed for new responses, M Turk older adults had higher recollection ratings (M = 1.99, SD = 1.43) compared to in-person older adults (M = 1.36, SD = 0.81), t (57) = 2.07, p <0.04 (Fig 21).

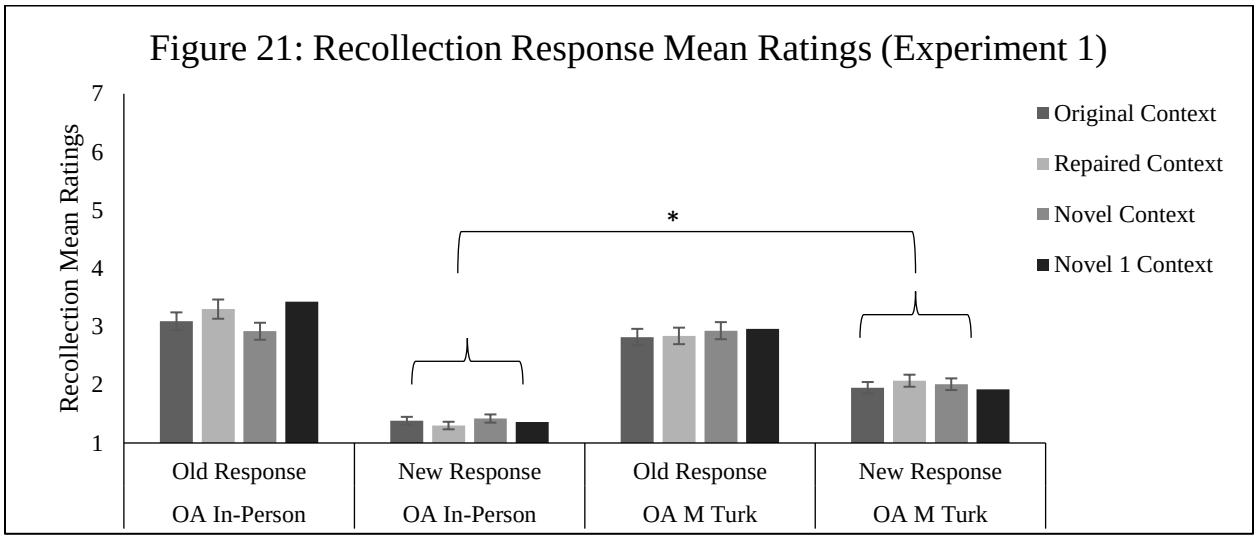


Figure 21: Mean recollection ratings for context between old response and new response across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of response by recruitment, F (1, 57) = 4.36, p<0.00

Covariates: After controlling for the total number of education and verbal score, the recruitment effect was still significant F (1,55) = 7.56, p<0.00.

Overall results suggest for older adults, context does not play a role in subjective ratings of recollection for contextual scenes. This is also reflected in their objective memory such as their cued recall or hits-false alarm rates. Interestingly, we saw that M Turk samples seem to have higher subjective ratings of recollection when they say “new” for old/new judgment for context scenes compared to the in-person sample. We suspect one reason could be that M Turk samples may not be paying attention to adjust their ratings to their old/new judgments. We don’t

think memory illusion for lure familiarity due to context reinstatement could take place for M Turk sample since it would show a significant difference in ratings between old scenes versus new scenes, which was not the case.

Familiarity for Context Condition by Recruitment

Familiarity for Response (Old vs. New) for Context Condition by Recruitment. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings of the amount of feeling of "oldness" for context scenes differed in terms of recruitment and context conditions. In other words, does participants pay high level of attention to context (in turn have higher feeling of oldness for "Old" scenes) and how does it differ for in-person and M Turk sample of older adults? To understand this question, we conducted a 4 (Context: Old, Repaired, Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Experiment: In-Person vs. M Turk) mixed ANOVA with context and response as a within-subject and recruitment as a between-subject factor to investigate the effect of context conditions, age and response type during old/new judgment phase on familiarity ratings. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as "Old" or incorrectly endorsed as "New."

Main Effects: There was a significant main effect of response F (1, 57) = 78.28, p <0.00 and a paired sample revealed old responses (M = 3.08, SD = 1.46) had higher ratings than new responses (M = 1.42, SD = 0.93), t (58) = 8.67, p <0.00.

Interaction: Results revealed a significant interaction of response by recruitment, F (1,57) = 1.02, p = 0.05. An independent sample revealed for new responses, M Turk (M = 1.72, SD =

1.21) sample had higher familiarity ratings compared to in-person sample (*M* = 1.13, *SD* = 0.71), *t* (57) = 2.30, *p* = 0.03 (Fig 22).

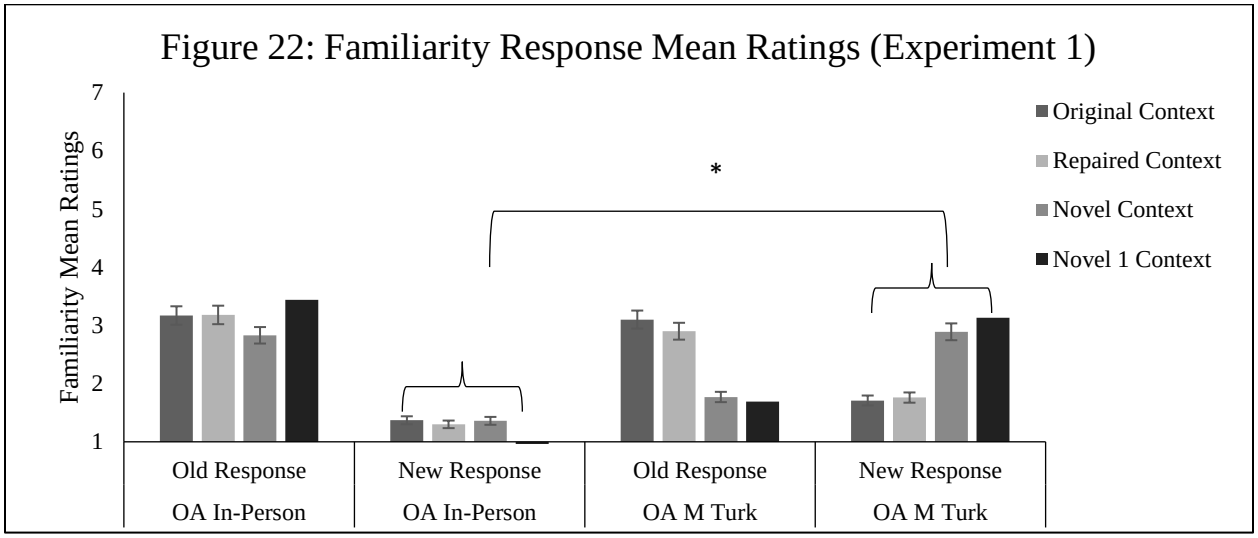


Figure 22: Mean familiarity ratings for context between old response and new response across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of response by recruitment, *F* (1, 57) = 1.02, *p*<0.00 and context by response *F* (3, 171) = 4.03, *p*<0.00. The latter was not significant after controlling for covariates.

There was also a significant 2-way interaction with context by response *F* (3, 171) = 4.03, *p*<0.00. To understand the interaction further, we first did two separate 4 (Context: Old, Repaired, Novel, Novel 1) x 2 (Recruitment: In-person vs. M Turk) mixed ANOVA with context as a within measure and recruitment as a between measure on Old responses and New responses for each test respectively to see whether ratings differ by response type across context and recruitment sample.

For old responses, we found no significant effects. For new responses, Mauchly's test revealed assumption of sphericity has been violated, χ^2 = (5) = 34.56, *p* < 0.00, ε = 0.70 (<0.75) therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. There was a main effect of context condition, *F* (2.10, 121.77) = 9.87, *p*<0.00. A paired sample t-

test revealed three significant differences. For new responses, familiarity ratings for original context (M = 1.54, SD = 1.11), repaired context (M = 1.52, SD = 1.20) and novel context (M = 1.53, SD = 1.05) in cued recall were greater compared to novel context exclusive to the condition (M = 1.09, SD = 1.13), t (58) = 3.79, p<0.00, t (58) = 3.70, p<0.00 and t (58) = 3.56, p<0.00.

Covariates: After controlling for the number of education and verbal score, only the recruitment effect was still significant, F (1, 55) = 6.15, p = 0.02.

Overall results suggest even though the context may not be accounted for their objective memory such as at cued recall, discrimination scores or hits for context scenes, older adults may still take context into account implicitly for their subjective ratings of familiarity, specifically having higher familiarity for scenes they have seen before at cued recall compared with novel scenes. Interestingly, replicating recollection ratings for scenes, we also saw that M Turk samples seem to have higher subjective ratings of familiarity when they say “new” for old/new judgment for context scenes compared to the in-person sample. We suspect similar low attention and in turn, poor calibration for new responses to ratings may have caused this pattern for M Turk.

Discussion for Experiment 1

In Experiment 1 our main aim was to investigate the impact of context reinstatement on the feeling of knowing ratings in younger and older adults. We found neither age differences nor context scene influence how participants make their FOK predictions. Although through our secondary analyses, we found an absence of response effects in the whole sample could have been influenced by our recruitment differences within older adults. M Turk sample seems to be well-calibrated on their predictions for recalled items while in-person sample seems to well-

calibrated for recalled items. One reason for this difference could be the differences in response ratings were overshadowed when older groups were combined and compared with younger adults. While when comparing younger and older adults, we may suggest an age equivalence for FOK predictions, results for response effects should be considered with caution and question generalizability in the whole population. We also did not find age or context effects to be influencing their objective memory such as cued recall or discrimination scores. However, participants, in general, may have attended to context, even if it was irrelevant to the target, to make some form of weak context memory signal and pair-context associations, as shown through objective memory for context scenes and subjective ratings for context-pair ratings.

We found for objective memory of context scenes, while younger adults seem to benefit from old context scenes explicitly, older adults, in general, do not regard context scenes for their memory, at least explicitly. Younger adults had higher accuracies than older adults at recognizing old scenes they have seen before (either original or repaired scene) shown through their hit rates of old scenes compared to novel scenes at cued recall. Older adults, in general, do not seem to hold context information into account whether they were recruited in-person or from M Turk, as shown by the absence of recruitment interaction in the recruitment analysis except cued recall.

For subjective ratings, we found age and context influence how much detail participants believe they can recall and whether they feel the pair was old. We found when compared with younger adults, older adults had higher ratings in recollecting more details for the pairs they think were old in the old/new judgment phase. We think the context effect pattern may have been confounded by sample differences within older adults. In our recruitment analysis for older

adults, we found a linear trend where ratings were higher for pairs in old scenes compared to new scenes, but only for pairs that were thought of as “new.” This pattern for new responses may be explained by the context-distortion hypothesis which explains reinstating context can also make lures, in our case perceptually similar lures to target, seem old, due to weak context-pair association made earlier. Specifically, we can expect this pattern to emerge in older adults as memory illusions are often associated with age deficiency in cognition.

For familiarity ratings, we found no effects for age, but we found effects for context. Results showed regardless of age, participants have higher subjective ratings of familiarity for lures shown in the original context compared to a novel context. We think several factors can influence such patterns. Considering this pattern was not replicated in the recollection item type for age groups, we think reinstating context to enhance memory illusions for lures may have had a bigger effect on familiarity for all participants, which generally does not focus on details but more on the gist of the object, thus easier to distort compared to recollection.

For subjective memory for context scenes, we found while context does not influence older adults’ subjective ratings, context scenes do seem to influence younger adults. We found when younger adults identify the background scenes as “new”, they seem to give higher recollection ratings for novel background scenes compared to an original background scene. One reason could be younger adults' subjective ratings of recollection seem to enhance whenever they try to identify the scene as old or new. Perhaps trying to recall some part of scenery increases their subjective feelings of recollection, even if it is incorrect. Memory illusion due to context reinstatement would not influence only for scenes since first, at this stage there was no presence of pairs to associate the scenes with. Second, these were novel scenes never seen before in the

early trials, thus not a reinstatement condition. The other reason could be younger adults simply did not calibrate their responses for old/new judgment with their subsequent recollection ratings.

It might also hint that recollection ratings could be influenced more by younger adults rather than older adults. The effects could be shown through our familiarity ratings for context scenes where instead of younger adults, we saw older adults having higher ratings in a similar pattern. Specifically, it seems to suggest context types come into play when older adults have low subjective feelings of familiarity and for new scenes, a similar pattern we see in recollection for younger adults. We think the effects for older adults could have been also influenced by recruitment differences, especially by the M Turk sample since we saw the ratings for new scenes reflected for the M Turk sample but not in-person sample.

In sum, from Experiment 1 we know age or context does not influence FOK prediction or objective memory. But for subjective memory, we see age and context effects, and it exists even when we control for covariates. We know older adults are most susceptible to memory illusions of lures as being target due to context reinstatement when they try to recollect details about the pairs in old scenes compared to new scenes. We found participants of all ages are more susceptible to these memory illusions when they try to recall vague details or feelings of "oldness" of the pairs. We also found that older adults are more prone to use familiarity for their subjective ratings compared to younger adults.

Experiment 2

We began with the idea in Experiment 1 that context might play a supportive role in FOK predictions and memory, at least in older adults. We found in contrast to what we thought earlier, context plays no supportive role in FOK predictions or objective memory. Therefore, in Experiment 2, we decided to examine whether presenting context as a distracting object would improve FOK predictions and memory at least for the pairs alone. Our setting mimicked that of a distraction task where instructions are to ignore the distractor. This is important to dissociate from dual attention or divided attention tasks since, in these tasks, participants are expected to pay attention to both the information and perform two concurrent tasks.

In Experiment 2, we examined whether the common FOK miscalibration in older adults could be reduced if older adults were explicitly directed to pay attention to the picture pairs instead of background scenes. Previous research has shown in general older adults have less memory and FOK accuracy than younger adults. While previous studies have focused on explicitly directing older adults to pay attention to any information related to the target, here we took a slightly different approach and asked older adults to instead focus their attention only towards the target, rather than surrounding information. Older adults, in general, have low suppression ability and poor attention, which in turn may make them pay attention more to the background scenes compared to the actual pairs they need to remember. From Experiment 1 that had no instructions, we know that context does not play a supportive role for FOK predictions when comparing age differences between younger and older adults. Older adults (and younger adults) do not take context into account when they either make their FOK predictions or on their objective memory such as their cued recall performance or their discrimination ability between

targets and lures. Both younger and older adults are age equivalent in these ratings. However, when it comes to subjective ratings, results hint that older adults may have taken in some form of partial context scenes and associated them with the pairs, as shown by their recollection-familiarity ratings.

Thus, in Experiment 2 our goal was to explore would their FOK predictions and objective memory improve if their attention was explicitly directed for target, thereby reducing distracting background? As an exploratory analysis, there were several directions for the results. But we hypothesized that explicit instruction to focus on the pairs might help older and younger adults to improve their FOK predictions for the pairs. Alternatively, due to age-deficiency in suppression ability, it might prompt older adults to pay more attention to context scenes compared to younger adults, thereby showing an age and context effect.

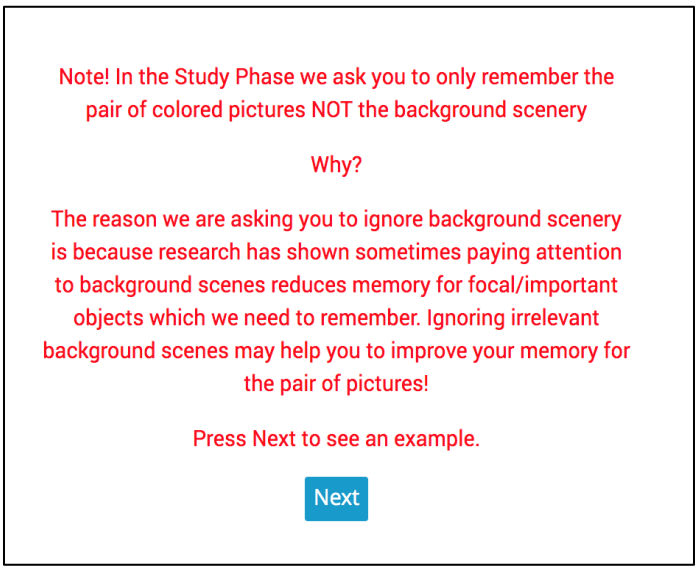
4. METHODOLOGY EXPERIMENT 2

Participants

The sample size and recruitment were the same as Experiment 1. Initially, we recruited 30 younger adults, 16 older adults in person, and 30 older adults via M Turk. As a reminder due to recruitment issues, we had a mix of the older adult sample from in-person and M Turk. One older participant was excluded from the M Turk sample due to scoring high on the GDS scale (GDS=14). Thus, for the analysis, we had 30 younger adults and 45 older adults (16 older adults recruited in person and 29 older adults recruited through M Turk). All participants were healthy (see Table 2 for demographic details).

Procedure

The procedure was same as Experiment 1 except before the study phase in each trial and practice phase, we show a screen where participants of all ages are asked to avoid background scenes for better memory in large font followed by visual example (see Fig 23)



23A: Text Instruction

23B: Visual Instruction

Figure 23: Experiment 2:

5. RESULTS EXPERIMENT 2

For each analysis, if there is a 3-way interaction, then it will be given priority and reported in detail while significant main effects or 2-way interaction will be mentioned briefly. In the absence of 3-way interaction, main effects and 2-way interaction will be reported in detail. This is done to avoid confusion with multiple analyses.

Covariate Analysis for Age

A bivariate correlation revealed most covariates correlated with one another (Table 12). An independent sample with Age as an IV and covariates as DVs were conducted to see which covariates differed across the two age groups (young vs. old) and help in deciding to choose the most important covariates for the analysis. We found across age, 3 covariates were significant: Younger adults had higher income (parent's income), t (73) = 4.77, p<0.00, lower verbal score t (73) = 8.51 p <0.00 and higher self-esteem t (73) = 3.41, p <0.00 compared to older adults. We created a composite score of the significant covariates as some of them correlated by combing their z-scores. This composite score was then entered for each analysis when we found an age effect without the covariates.

Table 12

Correlations for Experiment 2 by Age

		ED_ TOTAL	FATHER ED	MOTHER ED	INCOME	VERBAL SCORE	SELF ESTEEM	NCS
ED_TOTAL	Pearson Correlation	1.00	.13	-.11	.00	.05	-.43**	.24*
	Sig. (2-tailed)		.28	.34	.98	.68	.00	.04
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00
FATHERED	Pearson Correlation	.13	1.00	.24*	.34**	.06	.02	.07
	Sig. (2-tailed)	.28		.04	.00	.60	.90	.58
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00
MOTHERED	Pearson Correlation	-.11	.24*	1.00	.06	.07	.19	.22
	Sig. (2-tailed)	.34	.04		.60	.55	.10	.05
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00
INCOME	Pearson Correlation	.00	.34**	.06	1.00	-.32**	.03	.02
	Sig. (2-tailed)	.98	.00	.60		.01	.83	.88
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00
VERBALSCORE	Pearson Correlation	.05	.06	.07	-.32**	1.00	-.27*	.25*
	Sig. (2-tailed)	.68	.60	.55	.01		.02	.03
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00
SELF-ESTEEM	Pearson Correlation	-.43**	.02	.19	.03	-.27*	1.00	-.05
	Sig. (2-tailed)	.00	.90	.10	.83	.02		.69
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00
NCS	Pearson Correlation	.24*	.07	.22	.02	.25*	-.05	1.00
	Sig. (2-tailed)	.04	.58	.05	.88	.03	.69	
	N	75.00	75.00	75.00	75.00	75.00	75.00	75.00

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Analysis by Age (Younger adults vs. All Older adults)

Feeling of Knowing (FOK) Mean Ratings by Age

We wanted to investigate whether metamemory, aka FOK mean ratings differed in terms of age and context conditions. To understand this question, we conducted a 2 (Response: Recall vs. Unrecall) x 2 (Age: Young vs. Old) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with response and context as within-subject and age as a between-subject factor (Fig 7).

Mauchly's test revealed assumption of sphericity has been violated for context by response type, χ^2 (2) = 10.80, ε = 0.88 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.88>0.75) p<0.00, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity.

Main effects: Results revealed a main effect of response type, F (1,73) = 71.74, p< 0.00 such that participants gave higher FOK ratings for recalled items (M = 51.47, SD = 30.93) compared to unrecalled pairs (M = 15.95, SD = 16.26), t (74) =8.74, p< 0.00.

Interaction: There was a response type by age interaction, F (1, 73) = 11.46, p<0.00. An independent sample with Age as an IV and response type as a DV was conducted and results revealed within recalled items only, older adults (M = 60.34, SD = 30.03) have higher FOK ratings compared to younger adults (M = 38.16, SD = 27.71), t (73) = 3.23, p<0.00 (Fig 24).

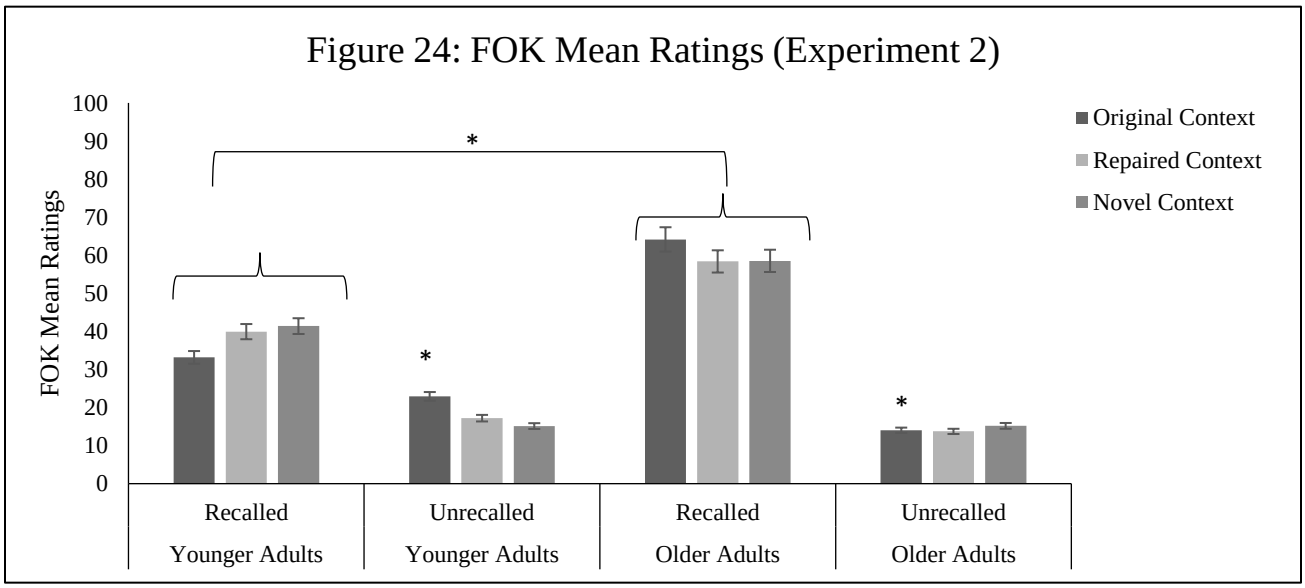


Figure 24: Mean Feeling of Knowing (FOK) ratings between recalled and unrecalled items across context and age. Scale range from 0-100. There was a significant 3-way interaction of context by response by age, F (2, 146) = 6.82, p<0.00.

The two-way interaction was followed by a three-way interaction for context by response type by age, F (2, 146) = 6.82, p<0.00. To understand the interaction, we conducted two separate analyses for Recall and Unrecalled responses.

For recalled responses, we found a significant Context by Age interaction, F (1.76, 128.15) = 4.56, p = 0.02. An independent t-test revealed within recalled responses, there were three significant effects. For all three types of context conditions, older adults had significantly higher response ratings for recalled pairs, compared to younger adults. Older adults (M = 64.14, SD = 27.56) had significantly higher FOK ratings compared to younger adults (M = 33.17, SD = 28.40) in original context, t (73) = 4.71, p<0.00. Older adults (M = 58.37, SD = 38.34) had

significantly higher FOK ratings compared to younger adults (M = 39.93, SD = 30.93) in repaired context, t (73) = 0.53, p = 0.02. Older adults (M = 58.52, SD = 33.02) also had significantly higher FOK ratings compared to younger adults (M = 44.38, SD = 33.19) in novel context, t (73) = 2.20, p = 0.03.

For unrecalled responses, we found a significant Context by Age interaction, F (2, 146) = 4.06, p = 0.02. An independent t-test revealed within unrecalled responses, there was a marginally significant effect for the original context, where younger adults had significantly higher FOK ratings compared to older adults, t (73) = 1.82, p<0.07.

Covariates: After controlling for covariates, the effect of age for the 2-way interaction and 3- way interaction were still significant, F (1, 72) = 10.69, p<0.00 and F (1.78, 128.17) = 7.82, p<0.00.

Overall results suggest age and background context play a role for FOK ratings depending on whether there are explicit instructions to avoid context. These effects exist even after controlling for covariates suggesting the effect is not due to third variable factors. In sum, it seems that older adults improve their FOK ratings with recalled items for all types of contextual scenes compared to younger adults. Although interpretation should be made with caution for unrecalled items due to marginal significance, the mean pattern seems to suggest younger adults are overconfident compared to older adults for pairs in the original context.

Cued Recall by Age

We wanted to investigate whether objective memory such as cued recall would differ with age and context conditions. To understand this question, we conducted a 2 (Age: Young vs.

Old) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with age as a between-subject factor and context as a within-subject factor.

Main effects: No significant main effects were found, F (2, 146) = 0.20, p = 0.82 (Table 13).

Table 13

Descriptive Statistics of Cued Recall Experiment 2 by Age

	Age	Mean	Std. Deviation	N
Recall_original	YA	.44	.29	30.00
	OA	.46	.32	45.00
	Total	.45	.30	75.00
Recall_repaired	YA	.44	.25	30.00
	OA	.44	.32	45.00
	Total	.44	.29	75.00
Recall_novel	YA	.46	.27	30.00
	OA	.42	.33	45.00
	Total	.44	.31	75.00

Interaction: No significant interactions were found context by age, F (2, 144) = 0.62, p = 0.54.

Overall results suggest older and younger adults are equivalent when making cued recall attempts and context condition has no effects when making such attempts.

Old/New Judgment by Age

We also measured another kind of objective memory besides cued recall—participant's ability to discriminate between target pairs from lure pairs that were perceptually similar to target pairs. The purpose was to explore whether sample differences played a role in recognizing target pairs better compared to lure pairs across context conditions in the cued recall phase.

Thus, two measures were obtained from old/new judgment: hit and false alarms. In the picture pair phase, hit rates were calculated as the proportion of correctly selecting “old” for the original cue-original target pair. In context only condition phase, hit rates were calculated as the proportion of correctly selecting “old” for original, repaired, and novel contexts which they have seen at cued recall. In the picture pair phase, false alarms were calculated as the proportion of incorrectly selecting “old” for the cue-lure target pair. In context only condition phase, a false alarm was calculated as the proportion of incorrectly selecting “old” for the new novel context exclusive at the context only condition, which participants have never seen before at cued recall.

After each measure was collected, for the picture pair phase, we calculation a discrimination score (Hits-FA) for each context condition—original, repaired, and novel contexts, reported below. We did not calculate discrimination score for context condition only but reported hits and false alarms separately since hits and false alarms did not have the same context conditions as novel context exclusive to the context only condition phase was not presented to participants before.

Discrimination Score (Hits-False Alarms) for Picture Pairs by Age. A 2 (Age: Young vs. Old) x 2 Context at Cued-Recall Test (Original, Repaired, Novel) mixed ANOVA with age as a between-subject factor and context as a within-subject factor was conducted to test the effect of context and age on participant’s discrimination ability between target vs. lure pairs across age and context conditions. After an analysis with outliers, 4 younger adults were excluded due to extremely high outliers for false alarms (falls above the 75th percentile of interquartile range). These outliers were only excluded for this analysis as for the rest of the analysis, they were not outliers.

Main effects: No significant effects were found, F (2, 138) = 0.19, p = 0.83 (Table 14).

Table 14

Group Statistics for Discrimination scores Experiment 2 by Age

	Age	N	Mean	Std. Deviation	Std. Error Mean
Pair_original	YA	26.00	.31	.43	.09
	OA	45.00	.47	.36	.05
Pair_repaired	YA	26.00	.35	.37	.07
	OA	45.00	.40	.35	.05
Pair_novel	YA	26.00	.26	.64	.13
	OA	45.00	.45	.30	.04

Interaction: No significant effects were found, context by age, F (2, 138) = 0.85, p = 0.43.

Overall results from discrimination scores from picture pairs seem to suggest background type and age does not affect the ability to distinguish between target and lures.

Hit for Context Condition Only by Age. A 2 (Age: Young vs. Old) x 3 Context at Cued-Recall Test (Original, Repaired, Novel) mixed ANOVA with age as a between-subject factor and context as a within-subject factor was conducted to test the effect of context and age on hit rates of background contexts only.

Main effects: No significant effects were found, F (2, 146) = 1.39, p = 0.25 (Table 15).

Table 15

Descriptive Statistics for Hits Context Condition Experiment 2 by Age

	Age	Mean	Std. Deviation	N
Context_Hits_original	YA	.54	.33	30.00
	OA	.29	.25	45.00
	Total	.39	.31	75.00
Context_Hits_repaired	YA	.52	.24	30.00
	OA	.30	.23	45.00
	Total	.39	.26	75.00
Context_Hits_novel	YA	.45	.25	30.00
	OA	.28	.28	45.00
	Total	.35	.28	75.00

Interaction: No significant interactions were found, context by age, F (2, 146) = 0.78, p = 0.46.

False Alarm for Context Condition Only by Age. An independent t-test was conducted with age as a between-group as independent variable and ratings of false alarms for novel context exclusive to the context condition as the dependent variable. These novel contexts were exclusive in the context conditions only, meaning participants have not seen these novel contexts before at the study or test phase. We wanted to see the FA rate (saying “old” to a context never shown before, which in this case would be a lure).

Main effects: No significant effects were found, t (73) = 1.00, p = 0.32 (Table 16).

Table 16

Group Statistics for FA Context Condition Experiment 2 by Age

	Age	N	Mean	Std. Deviation	Std. Error Mean
Context_FA	YA	30.00	.16	.12	.02
	OA	45.00	.13	.14	.02

Interaction: Not applicable.

Overall results suggest background has no influence on memory for context scenes between younger and older adults. One reason could be that participants may form context-pairs associations in both directions, such as context scenes that may inform pairs, pairs, in turn, can also inform context scenes. Thus, in a condition without pairs, perhaps there is a lack of memory for context scenes.

Recollection for Picture Pairs by Age

Recollection for Items (Target vs. Lures) for Picture Pairs by Age. We wanted to investigate whether subjective ratings of the number of details recollected differed in terms of age and context conditions for target pairs. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Pair: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and pair as a within-subject and age as a between-subject factor to investigate the effect of context conditions, age and pair type on recollection ratings. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether recollection differs depending on whether it was a target or a lure, in addition to context and age conditions.

Main effects: No significant effects were found, context F (2, 73) = 1.04, p = 0.36, item, F (1,73) = 1. 27, p = 0.26 (Table 17).

Table 17

Descriptive Statistics for Recollection Item Pairs Experiment 2 by Age

	Age	Mean	Std. Deviation	N
Recollection_Target_Original	YA	2.84	1.11	30.00
	OA	3.19	1.00	45.00
	Total	3.05	1.05	75.00
Recollection_Lure_Original	YA	2.98	1.44	30.00
	OA	2.56	1.58	45.00
	Total	2.73	1.53	75.00
Recollection_Target_Repaired	YA	2.97	1.09	30.00
	OA	3.19	.92	45.00
	Total	3.10	.99	75.00
Recollection_Lure_Repaired	YA	3.06	1.35	30.00
	OA	2.94	1.46	45.00
	Total	2.99	1.41	75.00
Recollection_Target_Novel	YA	2.90	1.32	30.00
	OA	3.10	.96	45.00
	Total	3.02	1.12	75.00
Recollection_Lure_Novel	YA	2.91	1.42	30.00
	OA	2.95	1.47	45.00
	Total	2.93	1.44	75.00

Interaction: No significant effects were found context by age, F (2, 146) = 0.30, p = 0.74, item type by age F (1, 73) = 3.37, p = 0.07, context by item type, F (2, 146) = 0.41, p = 0.67, context by item by age, F (2, 146) = 0.96, p = 0.39.

Overall, results suggest context type or age does not influence subjective ratings for the recollection of intact picture pairs or lure picture pairs. In other words, the result suggests an age

equivalence for subjective memory, specifically recollection ratings for picture pairs without any influence of background image.

Recollection for Responses (Old vs. New) for Picture pairs by Age. We also wanted to investigate whether subjective ratings of the number of details recollected differed in terms of age and context conditions for items participants deemed as "Old", in other words, pairs they think they have seen before. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Response: Old vs. New) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor was conducted to investigate the effect of context conditions, age and response type during old/new judgment phase on recollection ratings. This analysis considered whether the targets or lures were correctly endorsed as "Old" or "New".

Main effects: There was a main effect of response, F (1, 73) = 50.89, $p < 0.00$ and a paired sample t-test showed old responses (M =3.75, SD = 1.33) had higher ratings than new responses (M =2.19, SD = 1.24), t (74) = 7.40, p <0.00.

Interaction: There was a significant 2-way interaction of context by response, F (2, 146) = 3.29, p = 0.04. This was followed by a significant 3-way interaction between context by response by age F (2,146) = 3.36, p = 0.04 (Fig 25).

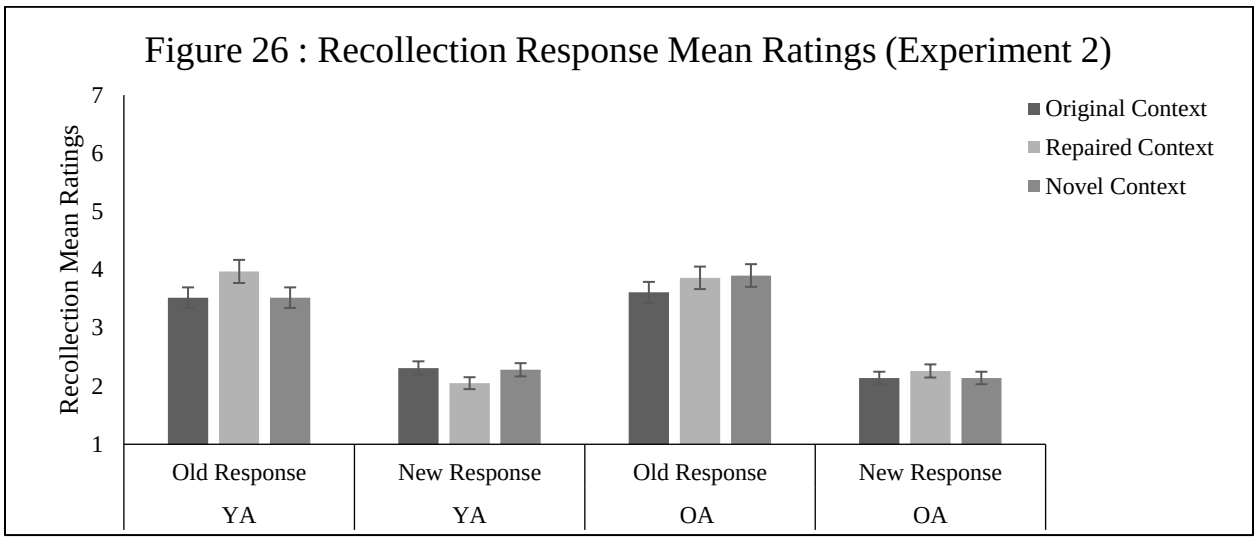


Figure 25: Mean recollection ratings for pairs between old response and new response across context and age. Scale range from 1-7. There was a significant 2-way interaction of context by response, F (2, 146) = 3.29, p = 0.04 and a significant 3-way interaction between context by response by age F (2,146) = 3.36, p = 0.04. The latter was not significant after controlling for covariates.

To understand a 3-way interaction, we did two 3 (Context: Old, Repaired, Novel) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and age a between measure on Old response and New response as the dependent variable for each test respectively.

For old responses, we found a main effect of context. A paired sample t-test revealed for old responses, recollection ratings for repaired context (M = 3.92, SD = 1.55) were significantly higher compared to original context (M = 3.57, SD = 1.40), t (74) = 2.30, p = 0.02. There were no age effects. For new items, the mixed ANOVA showed no significant effects, so no further analysis was conducted.

Covariates: The effect of age from the 3-way interaction was not significant after controlling for covariates, F (2, 144) = 2.87, p = 0.06.

Results seem to suggest that recollection ratings for response type are mainly driven by old responses ratings for original and repaired contexts and not age or context conditions after controlling for covariates.

Familiarity for Picture Pairs by Age

Familiarity for Items (Intact Target vs. Lures) for Picture pairs by Age. We wanted to investigate whether subjective ratings of the feeling of "oldness" differed in terms of age and context conditions for target pairs. To understand this question, we conducted a 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and item as a within-subject and age as a between-subject factor to investigate the effect of context conditions, age and item type on familiarity ratings. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether familiarity differs depending on whether it was a target or a lure, in addition to context and age conditions.

Main effects: There were no significant effects, context, F (2, 146) = 0.99, p = 0.38, item, F (1, 73) = 0.56, p = 0.46.

Interaction: There was a significant 2-way interaction for context by age for picture pairs, F (2, 146) = 3.44, p= 0.03, item by age F (1, 73) = 4.46, p = 0.04 followed by a significant 3-way interaction of context by item by age F (2, 146) = 5.40, p<0.00 (Fig 26).

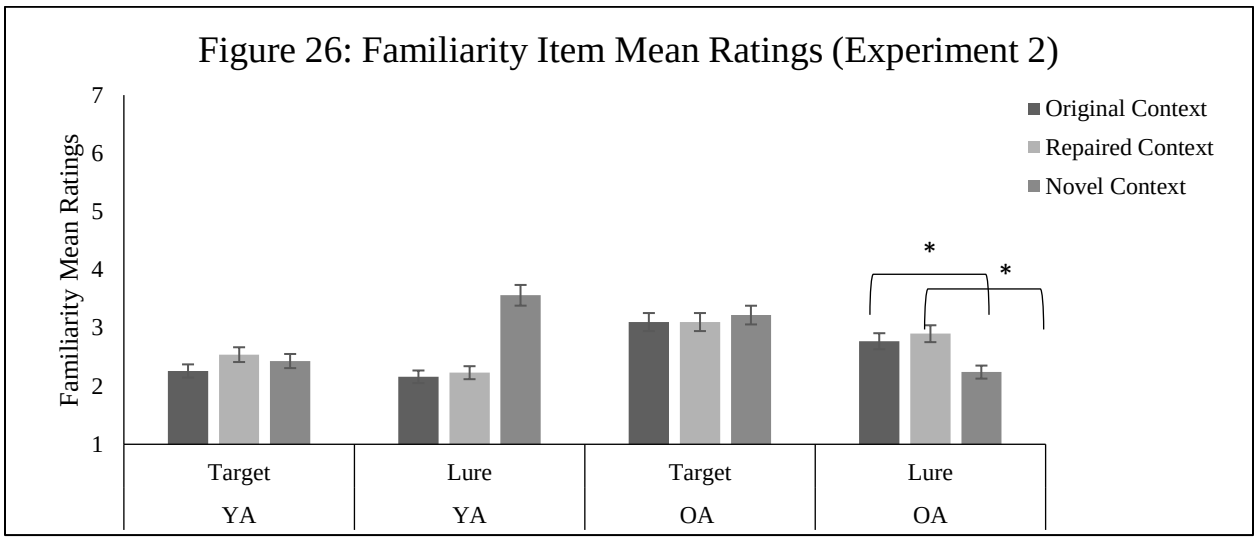


Figure 26: Mean familiarity ratings for pairs between target and lures across context and age. Scale range from 1-7. There was a significant 3-way interaction of context by age by item, F (2, 146) = 5.40, p<0.00

To understand the 3-way interaction, we first did two separate 3 (Context: Old, Repaired, Novel) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and age a between measure two separate ANOVA for intact item and for lure item to look for age effects for intact versus lure items separately.

For intact target items, there were no significant effects. For lures, there was a significant interaction of context by age, F (2, 146) = 4.57, p = 0.01 and to understand the interaction first an independent sample t-test showed older adults had higher familiarity ratings compared to younger adults for lure pairs in repaired context, t (73) = 2.11, p = 0.04. Then we split the file with age groups. There were no significant effects on younger adults. For older adults, there was a significant effect of context and a paired sample t-test across lures for older adults only revealed familiarity ratings for pairs revealed two effects. Ratings were greater for pairs in

original (M = 2.78, SD = 1.50) and repaired context (M = 2.90, SD = 1.40) compared to novel context (M = 2.24, SD = 1.24), t (44) = 2.26, p = 0.02, t (44) = 2.91, p<0.00

Covariates: The effect of Age was still significant after controlling for covariates, 2-way interaction, F (1, 72) = 4.18, p = 0.05 and 3-way interaction, F (2, 144) = 5.00, p<0.00.

Overall results suggest for subjective familiarity ratings of item types, there are no age differences for intact items. But for lures, older adults seem to mistakenly have higher familiarity compared to younger adults, but the inaccuracies in age seem to be only for lures in repaired context. Additionally, within older adults, it was found, an older adult has higher ratings for lures in original and repaired context compared to a novel context. Results seem to imply original and repaired context are held equally as "old scenes" when judging familiarity for lures by older adults. Also, older adults' poor binding ability may confuse them to think they have seen the perceptually similar lures in a previously seen context during context reinstatement effects. Overall results may hint that participants, especially older adults, in general, continue to make the association of context-pairs during the study phase even with explicit instructions to avoid context. Even with their poor attention to details, those partial associations made by older adults are strong enough to influence their later subjective memory.

Familiarity for Responses (Old vs. New) for Picture pairs by Age. We wanted to investigate whether subjective ratings of the feeling of "oldness" differed in terms of age and context conditions for items participants deemed as "Old", in other words, they think they have seen before. A 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor was conducted to investigate the effect of context conditions, age and

response type on familiarity ratings. This analysis considered whether the targets or lures were correctly endorsed as “Old” or “New”.

Mauchly’s test revealed assumption of sphericity has been violated for context, χ^2 (2) = 57.88, ε = 0.64 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.64<0.75) *p* = 0.03 and for context by item χ^2 (2) = 60.93, *ε* = 0.64 ($\varepsilon_{\text{Greenhouse Geisser}}$ = 0.64<0.75), *p* <0.00, therefore degrees of freedom were corrected using Greenhouse Geisser estimates of sphericity.

Main effects: There was a main effect of context, *F* (1.29, 94.05) = 4.20, *p* <0.03 with a paired sample t-test showing responses in novel context (*M* = 3.14, *SD* = 1.84) were greater compared to original context (*M* = 2.71, *SD* = 0.94), *t* (74) = 1.96, *p* <0.05.

Interaction: There was a significant 2-way interaction for context by response *F* (1.27, 92.94) = 4.52, *p* <0.01. The 2-way interaction was followed by a significant 3-way interaction of context by response by age, *F* (1.27, 92.94) = 7.68, *p*<0.00 (Fig 27).

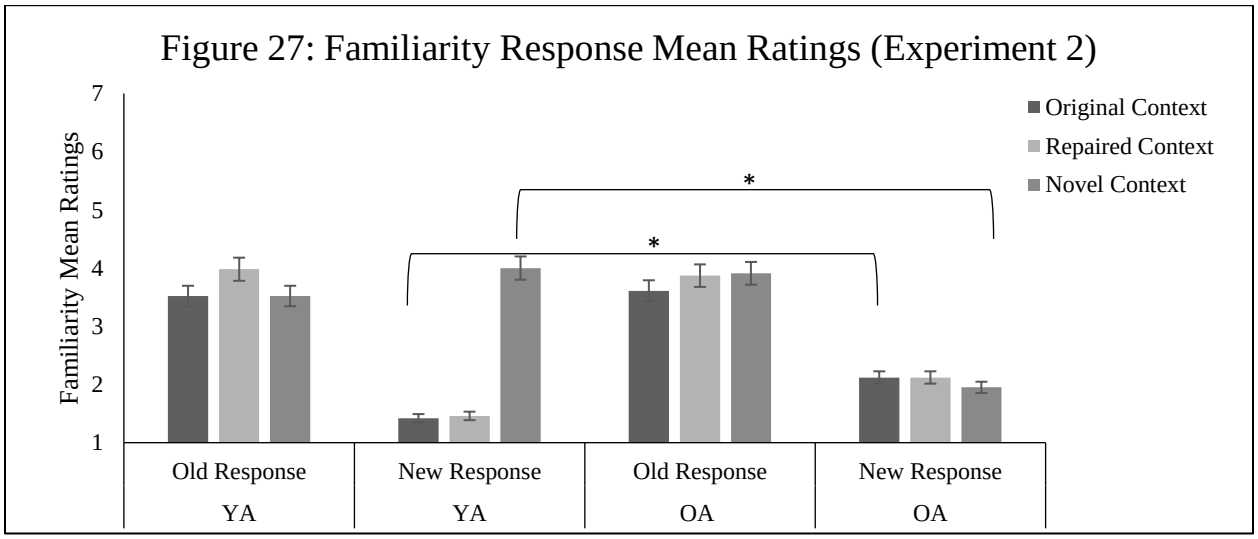


Figure 27: Mean familiarity ratings for pairs between old response and new response across context and age. Scale range from 1-7. There was a significant 3-way interaction of context by age by response, *F* (1.27, 92.94) = 7.68, *p*<0.00

To understand this interaction, we first did two separate 3 (Context: Old, Repaired, Novel) x 2 (Age: Young, Old) mixed ANOVA with context as a within measure and age a between measure on Old items and on New items to look whether ratings differ in old vs. new responses between younger and older adults.

For old responses, there was a significant main effect of context, F (2, 146) = 3.47, p = 0.03. A paired sample t-test for old responses showed ratings for pairs in repaired context were significantly higher than the original context, t (74) = 2.30, p = 0.02.

For new responses, there was a significant 2-way interaction for context by age, F (2, 146) = 6.39, p<0.00. To understand the interaction, first we looked for age effects and an independent sample t-test revealed older adults had higher ratings for new responses (M = 2.12, SD = 1.34) compared to younger adults (M = 1.42, SD = 0.99) in original context, t (73) = 2.48, p = 0.02. Older adults (M = 2.10, SD = 1.28) had higher ratings for new responses compared to younger adults (M = 1.46, SD = 0.13) in repaired context t (73) = 2.45, p = 0.03. Then we split the file by Age groups and redid the ANOVA for new response simple effects. Only younger adults had a main effect of context (2, 58) = 3.82, p = 0.03. A paired sample t-test across new responses for younger adults revealed, ratings for familiarity were higher for novel context (M = 3.40, SD = 5.29) compared to original context (M = 1.41, SD = 0.99), t (29) = 2.00, p = 0.05.

Covariates: The effect of age was still significant after controlling for covariates, F (2, 144) = 6.99, p<0.00.

Overall, results suggest older adults may have been more prone to memory illusions for lure familiarity in reinstated context. It may also suggest just as older adults, younger adults may also have been influenced by instruction conditions. However, given the fact that ratings are for

new items and in a novel context, it may suggest younger adults may not be simply paying attention to calibrate their old/new judgment with the familiarity ratings.

Recollection for Context Only Condition by Age

As a reminder, all three conditions from cued recall were considered as "Old" in the context condition only phase since participants have seen it before in the cued recall test. At this stage, they were intermixed with a new novel context condition (termed as "Novel 1" in the analysis) which was only reserved for old/new judgments phase at context conditions. This new context was deemed as "novel" since participants have never seen these contexts before at the cued-recall test. Since in this phase, we only showed the black and white context scenes without pairs, analysis can only be done for recollection/familiarity for response type (Old vs. New) from the Old/New Judgment condition in the context condition (read Procedure section/Fig. 1 for details).

Recollection for Response (Old vs. New) for Context Condition by Age. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings of the number of details recollected differed in terms of age and context conditions for only context scenes. In other words, do participants pay a high level of attention to context (in turn have higher recollected details for "Old" scenes) and how does it differ for different age groups? A 4 (Context: Old vs. Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor was conducted to investigate the effect of context conditions, age and response type during old/new judgment phase on recollection ratings. This analysis considered whether the original, repaired, and novel

context is *seen at the cued-recall test* were correctly endorsed as “Old” or incorrectly endorsed as “New.”

Main effects: A main effect of response was found, F (1, 73) = 48.97, p <0.00 but a further t-test revealed no significant effects. Main effect of context was not significant F (3, 219) = 1.51, p = 0.21 (Table 18).

Table 18

Descriptive Statistics for Recollection Response Context Condition Experiment 2 by Age

	Age	Mean	Std. Deviation	N
Recollection_Originalcontext_Old	YA	3.29	2.01	30.00
	OA	2.41	2.19	45.00
	Total	2.76	2.15	75.00
Recollection_Originalcontext_New	YA	1.96	2.35	30.00
	OA	1.40	.91	45.00
	Total	1.63	1.65	75.00
Recollection_Repairedcontext_Old	YA	3.54	1.93	30.00
	OA	2.89	2.41	45.00
	Total	3.15	2.24	75.00
Recollection_Repairedcontext_New	YA	1.81	1.43	30.00
	OA	1.41	.81	45.00
	Total	1.57	1.11	75.00
Recollection_Novelcontext_Old	YA	3.33	1.90	30.00
	OA	2.50	2.36	45.00
	Total	2.83	2.21	75.00
Recollection_Novelcontext_New	YA	1.98	1.46	30.00
	OA	1.39	.85	45.00
	Total	1.62	1.16	75.00
Recollection_Novel1context_Old	YA	3.42	1.66	30.00
	OA	2.53	1.68	45.00
	Total	2.88	1.72	75.00
Recollection_Novel1context_New	YA	1.47	1.22	30.00
	OA	1.32	.78	45.00
	Total	1.38	.98	75.00

Interaction: No significant effects were found, context by age F (3, 219) = 0.31, p = 0.82, response by age F (1, 73) = 0.94, p = 0.34, context by response, F (3, 219) = 1.61, p = 0.19, context by response by age, F (3, 219) = 0.43, p = 0.74.

Comparing results from pair conditions, results suggest that our instructions to avoid context may have worked, at least moderately. While partial attention may have been given to

context to form a weak context-pair association (suggested from subjective ratings for the pairs), high attention and in turn strong memory was not made to form the subjective amount of details for the scenes.

Familiarity for Context Condition by Age

Familiarity for Response (Old vs. New) for Context Condition by Age. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings for the feeling of "oldness" for context scenes differed in terms of age and context conditions. In other words, do participants pay a high level of attention to context (in turn have a higher feeling of oldness for "Old" scenes) and how does it differ for different age groups? A 4 (Context: Old vs. Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor was conducted to investigate the effect of context conditions, age and response type during old/new judgment phase on familiarity ratings. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as "Old" or incorrectly endorsed as "New."

Main Effects: There was a significant main effect of response F (1, 73) = 16.16, p <0.00 and a paired sample revealed old responses (M = 3.47, SD = 4.81) had higher ratings than new responses (M = 1.35, SD = 0.88), t (74) = 3.74, p <0.00. Main effect of context was not significant, F (3, 219) = 2.14, p = 0.1 (Table 19).

Table 19

Descriptive Statistics for Familiarity Response Context Condition Experiment 2 by Age

	Age	Mean	Std. Deviation	N
Familiarity_Originalcontext_Old	YA	3.31	1.86	30.00
	OA	2.53	2.30	45.00
	Total	2.84	2.15	75.00
Familiarity_Originalcontext_New	YA	1.52	1.37	30.00
	OA	1.41	.95	45.00
	Total	1.46	1.13	75.00
Familiarity_Repairedcontext_Old	YA	8.73	28.81	30.00
	OA	2.92	2.43	45.00
	Total	5.24	18.36	75.00
Familiarity_Repairedcontext_New	YA	1.79	1.36	30.00
	OA	1.37	.76	45.00
	Total	1.54	1.06	75.00
Familiarity_Novelcontext_Old	YA	3.28	1.90	30.00
	OA	2.56	2.38	45.00
	Total	2.85	2.22	75.00
Familiarity_Novelcontext_New	YA	1.70	1.11	30.00
	OA	1.37	.91	45.00
	Total	1.50	1.00	75.00
Familiarity_Novel1context_Old	YA	3.42	1.63	30.00
	OA	2.64	1.69	45.00
	Total	2.95	1.70	75.00
Familiarity_Novel1context_New	YA	.84	1.04	30.00
	OA	.94	.97	45.00
	Total	.90	.99	75.00

Interaction: No significant interaction was found, context by age, F (3, 129) = 1.52, p = 0.21, response type by age F (1, 73) = 2.55, p = 0.12, context by response type, F (3, 219) = 1.49, p = 0.22, context by response by age, F (3, 219) = 1.19, p = 0.32.

Results suggest that context scenes that seemed old also seemed more familiar to participants compared to context scenes that seemed new, regardless of age or context conditions. This might also hint that our instructions to avoid context may have worked, at least moderately. While partial attention may have given to context to form a weak context-pair association (suggested from subjective ratings for the pairs), high attention and in turn strong memory was not made to form a subjective feeling of "oldness" for the scenes.

Analysis by Recruitment for Older Adults (In-person versus M Turk)

Demographics and Covariate Analysis for Recruitment

See Table 2 for demographics. An independent sample t-test was conducted to investigate whether M Turk older adult samples differed from the in-person older adult sample in terms of demographics. The tt-test revealed in-person sample had higher number of years in education level, t (43) = 4.04, p<0.00, higher income, t (43) = 1.99, p = 0.05, had low level of mother's education t (43) = 2.39, p = 0.02, had lower self-esteem score t (43) = 10.58, p<0.00. Due to the high collinearity of these measures a composite score was created as before. Similar analysis as before was conducted whenever there significant was a recruitment effect in the main analysis.

A similar analysis for FOK, Recall, Discrimination scores, Recollection-Familiarity measures was conducted as above replacing Age group with Recruitment group for older adults (In-person vs. M Turk) to investigate whether recruitment differences for older adults influenced experimental measures.

Feeling of Knowing (FOK) Mean Ratings by Recruitment

We wanted to investigate whether metamemory, aka FOK mean ratings differed in terms of recruitment and context conditions. A 2 (Response: Recall vs. Unrecall) x 2 (Recruitment: OA M Turk vs. OA in-person) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with response and context as within-subject and Recruitment as a between-subject factor was conducted to investigate where FOK mean ratings differed in terms of Recruitment and context conditions.

Main Effect: There was a main effect of response type F (1, 43) = 96.48, p<0.00 and a paired sample t-test revealed recalled items (M = 60.34, SD = 30.03) had higher FOK ratings than unrecalled items (M = 14.31, SD = 12.99), t (44) = 10.53, p<0.00.

Interaction: Results revealed a significant 2-way interaction of context by response type, F (2, 86) =3.52, p = 0.03. To understand the interaction two separate ANOVA was done for recalled and unrecalled items as before (Fig 28).

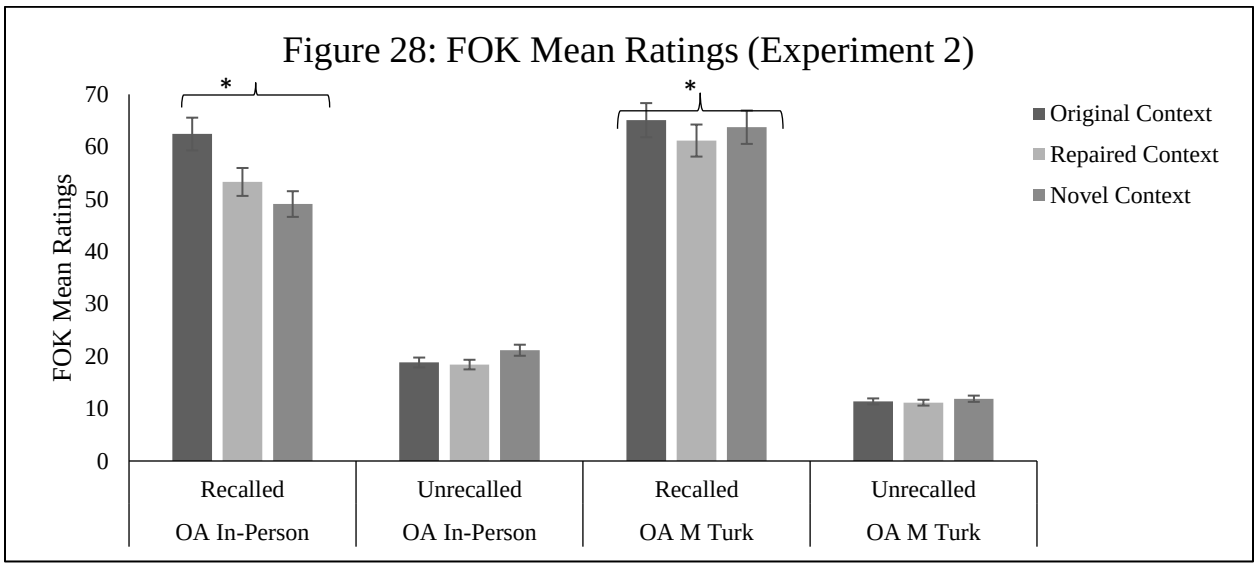


Figure 28: Mean Feeling of Knowing (FOK) ratings between recalled and unrecalled items across context and recruitment. Scale range from 0-100. There was a 2-way interaction of context by response, F (2, 86) = 3.52, p= 0.03.

For recalled items there was a main effect of context F (2, 86) = 1.94, p = 0.04 and a paired sample t-test revealed for recalled items, FOK ratings were marginally significantly higher for original context (M = 64. 14, SD = 27.56) compared to repaired context (M = 58.37, SD = 35.35), t (44) = 1.93, p = 0.06. There were no significant effects for unrecalled items or for novel context.

Overall results suggest there are no recruitment differences within the older adult sample to influence their FOK ratings. Older adults, in general, can estimate well for items they recalled compared to items they cannot at the moment. Although interaction significance is marginal and should be taken to caution for interpretation, older adults seem more confident for recalled items in original context compared to a repaired context implying they may have made some partial association with the pairs and the context at the study phase even with explicit instructions to avoid context.

Cued Recall for Pairs by Recruitment

We wanted to investigate whether objective memory such as cued recall would differ with recruitment and context conditions. A 2 (Recruitment: OA M Turk vs. OA in-person) x 3 (Context: Original vs. Repaired vs. Novel) mixed ANOVA with Recruitment as a between-subject factor and context as a within-subject factor was conducted to investigate the effect of context conditions on cued recall memory performance.

Main effects: No significant effects were found, F (2, 86) = 1.03, p = 0.36 (Table 20).

Table 20

Descriptive Statistics of Cued Recall Experiment 2 by Recruitment

Factors	Recruitment	Mean	Std. Deviation	N
Recall_original	In-person OA	.29	.23	16.00
	MturkOA	.55	.32	29.00
	Total	.46	.32	45.00
Recall_repaired	In-person OA	.32	.28	16.00
	MturkOA	.51	.32	29.00
	Total	.44	.32	45.00
Recall_novel	In-person OA	.24	.25	16.00
	MturkOA	.52	.33	29.00
	Total	.42	.33	45.00

Interaction: No significant effects were found, context by recruitment, F (2, 86) = 1.16, p = 0.32.

Overall results suggest a type of context or sample type does not influence cued recall for older adults. This is similar for the whole sample when older adults sample were combined and compared with younger adults. This may mean our results are generalizable, at least concerning this experiment as there is no recruitment difference.

Old/New Judgment for Pairs by Recruitment

Discrimination Score (Hits-False Alarms) for Picture Pairs by Recruitment. A 2 (Age: Young vs. Old) x 2 Context at Cued-Recall Test (Original, Repaired, Novel) mixed ANOVA with recruitment as a between-subject factor and context as a within-subject factor was conducted to test the effect of context and recruitment on participant's discrimination ability between target vs. lure pairs.

Main effects: No significant effects were found, F (2, 86) = 0.89, p = 0.41 (Table 21).

Table 21

Descriptive Statistics of Discrimination Scores Experiment 2 by Recruitment

Factors	Recruitment	Mean	Std. Deviation	N
Pairs_Original	In-person OA	.40	.34	16.00
	MturkOA	.51	.37	29.00
	Total	.47	.36	45.00
Pairs_Repaired	In-person OA	.36	.35	16.00
	MturkOA	.42	.36	29.00
	Total	.40	.35	45.00
Pairs_Novel	In-person OA	.42	.31	16.00
	MturkOA	.47	.30	29.00
	Total	.45	.30	45.00

Interaction: No significant effects were found, context by recruitment, F (2, 86) = 0.25, p = 0.78.

Overall results suggest for older adults exclusively, background type or recruitment does not influence memory discrimination for picture pairs or when older participants make errors in their memory discrimination.

Hits for Context Condition Only by Recruitment. A 2 (Recruitment: OA M Turk vs. OA in-person) x 2 Context (Original vs. Novel) mixed ANOVA with Recruitment as a between-subject factor and context as a within-subject factor was conducted to test the effect of context and age on hit rates of contexts.

Main effects: No significant effects were found, F (2, 86) = 0.36, p = 0.70 (Table 22).

Table 22

Descriptive Statistics for Hits for Context Condition Experiment 2 by Recruitment

Factors	Recruitment	Mean	Std. Deviation	N
Context_Hits_original	In-person OA	.20	.26	16.00
	MturkOA	.34	.24	29.00
	Total	.29	.25	45.00
Context_Hits_repaired	In-person OA	.25	.22	16.00
	MturkOA	.33	.24	29.00
	Total	.30	.23	45.00
Context_Hits_novel	In-person OA	.19	.25	16.00
	MturkOA	.33	.29	29.00
	Total	.28	.28	45.00

Interaction: No significant effects were found context by recruitment, F (2, 86) = 0.43, p = 0.65.

FA for Context Only Condition by Recruitment. An independent t-test was conducted with Recruitment as a between-group as independent variable and ratings of false alarms for original context only as a dependent variable since false alarms cannot be computed in a novel context.

Main effects: No significant effects were found, t (43) = 0.45, p = 0.65 (Table 23).

Table 23

Group Statistics for FA Context Condition Experiment 2 by Recruitment

	Recruitment	N	Mean	Std. Deviation	Std. Error Mean
Context_FA	In-person OA	16	.1122	.13753	.03438
	MturkOA	29	.1325	.14712	.02732

Interaction: Not applicable.

Overall results suggest within older adults, there is no effect of context or sample difference on discriminating between old or new scenes.

Recollection for Picture Pairs by Recruitment

Recollection for Items (Target vs. Lures) for Picture Pairs by Recruitment. We wanted to investigate whether subjective ratings of the amount of details recollected differed in terms of recruitment and context conditions for target pairs. A 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Recruitment: OA M Turk vs. OA in-person) mixed ANOVA with context and item as a within-subject and Recruitment as a between-subject factor was conducted to investigate the effect of context conditions, age and item type on recollection ratings. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether recollection differs depending on whether it was a target or a lure, in addition to context and age conditions.

Main effect: No significant effects were found, context F (2, 86) = 1.74, p =0.18 or item F (1, 43) = 3.19, p = 0.08 (Table 24).

Table 24

Descriptive Statistics for Recollection Item Pairs Experiment 2 by Recruitment

Factors	Recruitment	Mean	Std. Deviation	N
Recollection_Target_Original	In-person OA	3.28	.49	16.00
	MturkOA	3.13	1.19	29.00
	Total	3.19	1.00	45.00
Recollection_Lure_Original	In-person OA	2.55	1.20	16.00
	MturkOA	2.56	1.78	29.00
	Total	2.56	1.58	45.00
Recollection_Target_Repaired	In-person OA	3.12	.60	16.00
	MturkOA	3.22	1.06	29.00
	Total	3.19	.92	45.00
Recollection_Lure_Repaired	In-person OA	3.36	1.30	16.00
	MturkOA	2.71	1.51	29.00
	Total	2.94	1.46	45.00
Recollection_Target_Novel	In-person OA	3.51	.85	16.00
	MturkOA	2.87	.95	29.00
	Total	3.10	.96	45.00
Recollection_Lure_Novel	In-person OA	3.33	1.36	16.00
	MturkOA	2.74	1.51	29.00
	Total	2.95	1.47	45.00

Interaction: No significant effects were found, context by recruitment, F (2, 86) = 1.92, p = 0.15, item type by recruitment F (1, 43) = 1.50, p = 0.26, context by item type, F (2, 86) = 1.89, p = 0.16, context by item by recruitment, F (2, 86) = 1.37, p = 0.26.

Overall results suggest within older adults, there is no effect of context or sample difference on discriminating between old or new scenes, implying any significant effects we saw in the whole sample were not influenced by sample differences.

Recollection for Response (Old vs. New) for Picture Pairs by Recruitment. We also wanted to investigate whether subjective ratings of the number of details recollected differed in terms of recruitment and context conditions for items participants deemed as “Old”, in other

words, pairs they think they have seen before. A 3 (Context: Original vs. Repaired vs. Novel) X 2 (Response: Old vs. New) X 2 (Recruitment: OA M Turk vs. OA in-person) mixed ANOVA with context and response as a within-subject and Recruitment as a between-subject factor was conducted to investigate the effect of context conditions, Recruitment and response type during old/new judgment phase on recollection ratings. This analysis considered whether the targets or lures were correctly endorsed as “Old” or “New”.

Main Effect: There was a main effect of response type F (1, 43) = 247.80, p< 0.00 and a paired sample t-test revealed recollection ratings for old responses had higher ratings (M = 3.79, SD = 1.39) compared to new responses (M = 2.18, SD = 1.16), t (44) = 5.56, p<0.00.

Interaction: There was a significant 2-way interaction of response by recruitment, F (1, 43) = 176.96, p< 0. 00 (Fig 29).

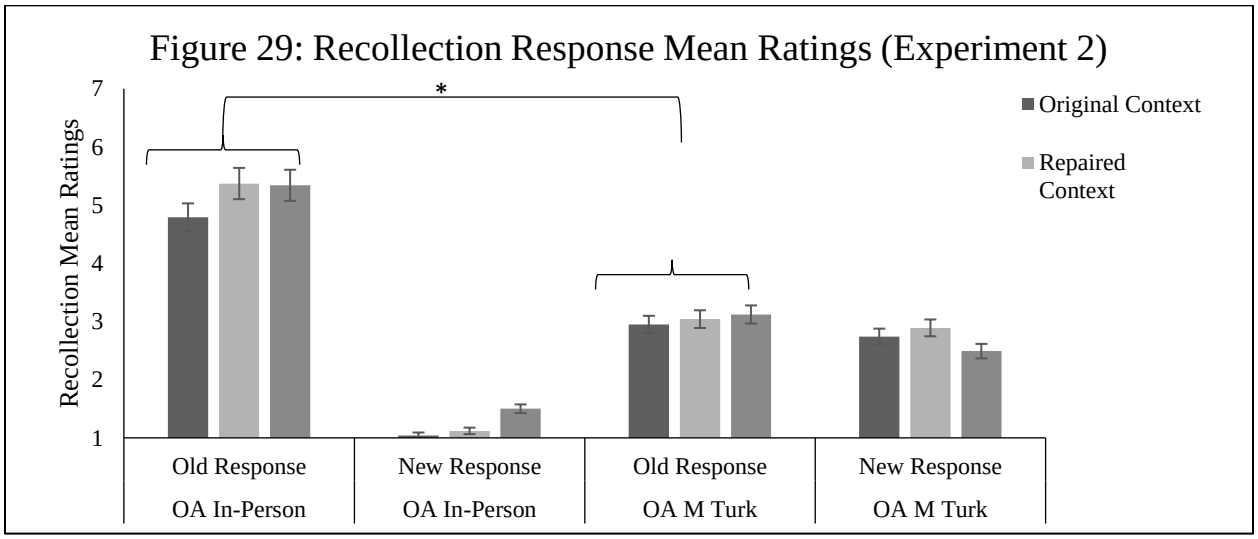


Figure 29: Mean recollection ratings for pairs between old response and new response across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of response by recruitment, F (21, 43) = 176.96, p<0.00 and a 3-way interaction of context by response by age F (2, 86) = 3.42, p<0.00. The latter was not significant after controlling for covariates

An independent sample t-test showed for old responses, in-person sample (*M*= 5. 17, *SD* = 0.85) had higher ratings compared to M Turk sample (*M*= 3.04, *SD* = 1.00), *t* (43) = 7.21, *p*<0.00. For new responses, M Turk sample (*M*= 2.71, *SD* = 0.87) had higher ratings compared to in-person sample (*M*= 1.22, *SD* = 1.01), *t* (43) = 5.19, *p*<0.00.

This was followed by a significant 3-way interaction for context by response type by recruitment, *F* (2, 86) = 3.43, *p*<0.04. To understand this interaction, first, a paired sample t-test revealed older adults in M Turk sample had higher ratings for new responses compared to the in-person sample in all three context condition, *t* (43) = 4.89, *p*<0.00, *t* (43) = 5.44, *p*<0.00, *t* (43) = 2.77, *p*<0.00. Then we did two separate mixed ANOVAs for old responses and new responses.

We found no significance for old responses. For new responses, we found there was a significant interaction of context by recruitment, *F* (2, 86) = 3.37, *p* = 0.04 and to understand the interaction we split the file by Recruitment groups and redid the ANOVA for new responses. No significant effects were found.

Covariates: The effect of recruitment was significant after controlling for covariates for the 2-way interaction, *F* (1, 42) = 157.52, *p*<0.00 but not for the 3-way interaction, F (2, 84) = 2.65, p = 0.08.

Overall results suggest context may not influence older adults in general. However, older adults in person have higher subjective ratings for old items. M Turk sample seems to inaccurately have higher subjective ratings of recollected details for new items. The 3-way interaction with context disappears after controlling for covariates suggesting the sample differences in ratings are mainly driven by response types between M Turk and In-person sample and not context scenes. Context effects only occur due to covariates. Previous research has

shown having higher education and in turn, higher-income seem to enhance memory accuracies in general. Thus, it may suggest having higher cognitive reserve factors may protect in-person samples with accurate calibration between old items in the old/new judgment and recollection ratings compared to M Turk sample. Additionally, the M Turk sample may also not be calibrating well due to a lack of attention for the whole study.

Familiarity for Picture Pairs by Recruitment

Familiarity for Items (Target vs. Lures) for Picture Pairs by Recruitment. We wanted to investigate whether subjective ratings of the feeling of "oldness" differed in terms of recruitment and context conditions for target pairs. A 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Recruitment: OA M Turk vs. OA in-person) mixed ANOVA with context and item as a within-subject and Recruitment as a between-subject factor was conducted to investigate the effect of context conditions, Recruitment and item type on familiarity ratings. This analysis did not consider whether the targets or lures were correctly endorsed. Rather we wanted to test whether familiarity differs depending on whether it was a target or a lure, in addition to context and recruitment conditions.

Main effects: There was a main effect of context, F (2, 86) = 3.28, p = 0.04. A paired sample t-test revealed across items, ratings for items in repaired context (M = 3.00, SD = 0.89) were significantly higher than novel context (M = 2.73, SD = 0.84), t (44) = 2.12, p = 0.04.

There was a main effect of item type F (1, 43) = 10.35, p<0.00 with intact items having significantly higher familiarity ratings (M = 3.70, SD = 1.13) compared to lure items (M = 2.55, SD = 1.03), t (44) = 5.52, p< 0.00.

Interaction: Results revealed a significant 2-way interaction of context by item, F (2, 86) = 7.13, p<0.00 followed by a significant 3-way interaction of context by item by recruitment, F (2, 86) = 3.72, p = 0.03 (Fig 30).

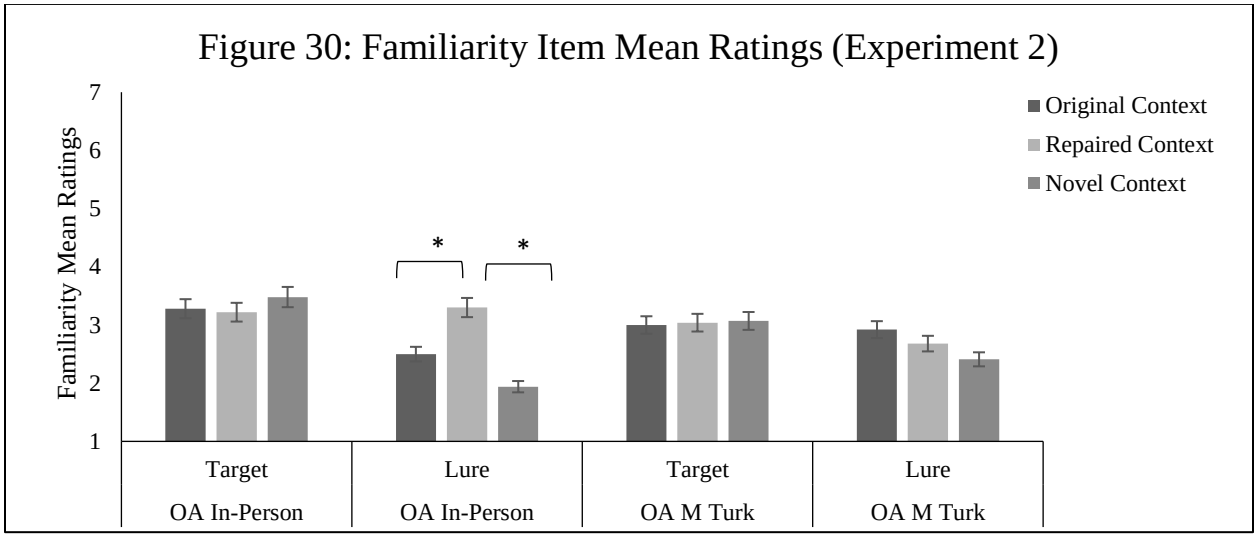


Figure 30: Mean familiarity ratings for pairs between target and lure across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of context by recruitment, F (2, 86) = 4.99, p < 0.00 and response by recruitment F (1, 43) = 121.26, p < 0.00.

To understand this interaction, we first did two separate mixed ANOVAs for intact and lure items as before.

We found for intact items there were no significant effects. For lure items we found a main effect of context, F (2, 86) = 6.58, p<0.00. A paired sample t-test revealed two significant effects. For lure items overall, familiarity ratings were higher for pairs in original context (M = 2.77, SD = 1.50) and repaired context (M = 2.90, SD = 1.40) compared to ratings for pairs in novel context (M = 2.24, SD = 1.25), t (44) =2.26, p = 0.03, t (44) = 2.91, p< 0.00.

There was also a significant effect of context by recruitment, F (2, 86) = 3.60, p = 0.03. Independent sample t-test scores lure showed no recruitment differences. So, to examine how

ratings differ between each in-person and M Turk individually, we split the file by Recruitment groups and redid the ANOVA for lures. We found for older adults in-person only, there was a main effect of context. A paired sample t-test revealed across lures for in-person groups, ratings for pairs in repaired context (M = 3.26, SD = 0.68) was higher compared to original context (M = 2.89, SD = 0.58), t (15) = 2.73, p = 0.02. Ratings were also significantly higher for repaired context compared to novel context (M = 2.17, SD = 0.20), t (15) = 2.73, p = 0.02.

Covariates: The effect of recruitment was still significant after controlling for covariates, F (2, 84) = 4.48, p = 0.01.

Overall results seem to suggest the interaction effects are mainly driven by in-person older adults for lures. Specifically, it may imply in-person older adults seem to value both original and repaired context as “old context” and due to some partial association made between pair and context at study scenes, their familiarity seems to be higher for these old contexts compared to a new context. Due to poor attention to detail, they mistakenly might presume perceptually similar lures to be the same as target pairs. Thus, old context paired with perceptually similar lures may seem more familiar to them compared to pairs in a new context. On the other hand, context or item type seems to not influence the M Turk sample.

Familiarity Response (Old vs. New) for Picture Pairs by Recruitment. We wanted to investigate whether subjective ratings of the feeling of “oldness” differed in terms of age and context conditions for items participants deemed as “Old”, in other words, they think they have seen before. A 3 (Context: Original vs. Repaired vs. Novel) X 2 (Item: Target vs. Lure) X 2 (Age: Young vs. Old) mixed ANOVA with context and response as a within-subject and age as a between-subject factor was conducted to investigate the effect of context conditions, age and

response type on familiarity ratings. This analysis considered whether the targets or lures were correctly endorsed as “Old” or “New”.

Main effect: There was a main effect of response, F (1, 43) = 193.56, p< 0.00. A paired sample t-test revealed old responses (M = 3.72, SD = 1.16) having higher ratings than new responses (M = 2.54, SD = 0.81), t (44) = 2.08, p<0.00

Interaction: Results revealed a significant interaction (Fig 31) of context by recruitment F (2, 86) = 4.99, p< 0.00 and an independent sample t-test revealed across response types, in-person sample (M = 3.37, SD = 0.70) had higher ratings compared to M Turk sample (M = 2.68, SD = 0.87) for pairs in novel context only, t (43) = 2.73, p<0.00.

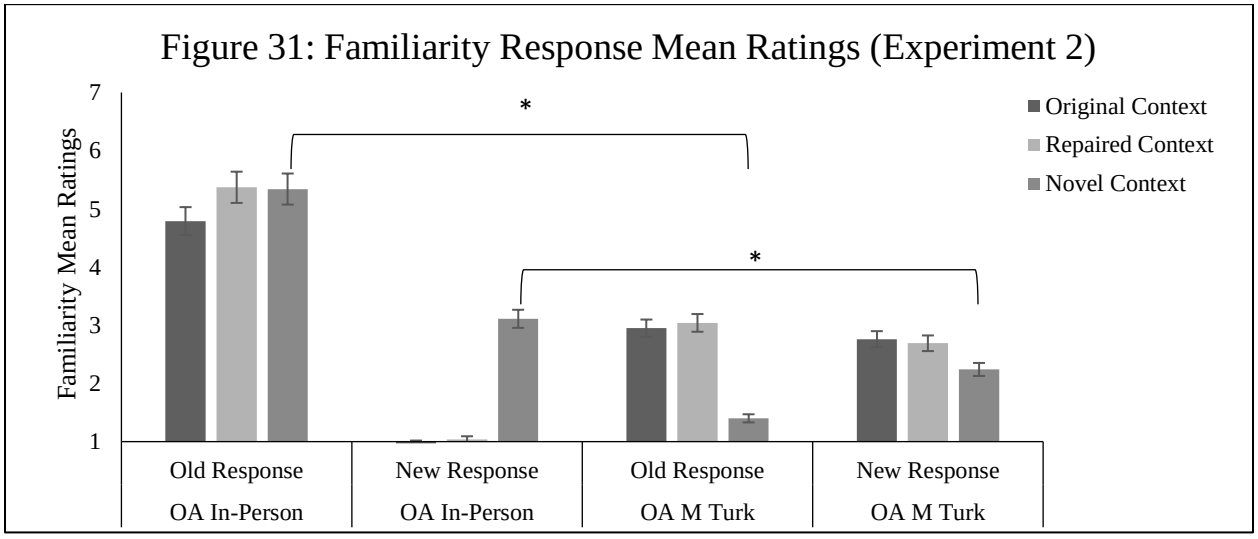


Figure 31: Mean familiarity ratings for pairs between old and new responses across context and recruitment. Scale range from 1-7. There was a significant 2-way interaction of context by recruitment, F (2, 86) = 4.99, p < 0.00 and response by recruitment F (1, 43) = 121.26, p < 0.00.

Results revealed a significant interaction (Fig 31) of response by recruitment F (1, 43) = 121.26, p< 0.00 and an independent sample t-test revealed for old responses only in-person

sample (M = 4.77, SD = 0.80) had higher ratings compared to M Turk sample (M = 3.14, SD = 0.99), t (43) = 6.08, p<0.00.

Covariates: The effect of recruitment was still significant after controlling for covariates, F (2, 84) = 6.27, p<0.00 and F (1, 42) = 108.42, p<0.00.

Overall results suggest across responses, in-person sample have higher familiarity for pairs in a novel context and old pairs compared to M Turk sample showing perhaps in-person sample have better calibration for their subjective ratings for items they think as "Old".

Recollection for Context Only Condition by Recruitment

Recollection for Response (Old vs. New) for Context Condition by Recruitment. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings of the number of details recollected differed in terms of recruitment and context conditions for only context scenes. In other words, do participants pay a high level of attention to context (in turn have higher recollected details for "Old" scenes) and how does it differ for in-person and M Turk sample of older adults? A 4 (Context: Old vs. Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Recruitment: In-person vs. M Turk) mixed ANOVA with context and response as a within-subject and Recruitment as a between-subject factor was conducted to investigate the effect of context conditions, recruitment and response type during old/new judgment phase on recollection ratings. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as "Old" or incorrectly endorsed as "New."

Main effects: There was a main effect of response, F (1, 43) = 20.99, p <0.00 and paired sample t-test revealed across response types, ratings for scenes with old responses (M = 2.58, SD = 1.81) have higher ratings than new (M = 1.38, SD = 0.80) responses, t (44) = 4.98, p <0.00.

Interaction: No significant effects were found, context by recruitment, F (3, 129) = 0.32, p = 0.81, response type by recruitment, F (1, 43) = 0.24, p = 0.63, context by response type, F (3, 129) = 0.92, p = 0.44, context by response by recruitment, F (3, 129) = 0.16, p = 0.93.

Overall results suggest for older adults, context or sample differences do not play a role in subjective ratings of recollection for contextual scenes. This is also reflected in their objective memory such as their cued recall or hits-false alarm rates. Perhaps older adults pay attention to the actuality of the instruction to avoid context scenes. Hence, while they do not explicitly pay attention to the context scenes, implicitly at least (hinted by their subjective ratings for the pairs) suggest they pay partial attention to the scenes and form instead of a vague context-pair association. Thus, their ratings for the feeling of "details" of context scenes are not influenced by context conditions or sample type. It might overall suggest that explicit instruction has helped older adults to avoid context scenes to form any strong subjective memory of details.

Familiarity for Context Only Condition by Recruitment

Familiarity for Response (Old vs. New) for Context Condition by Recruitment. The purpose was to separate context memory from context-pair associative memory. We wanted to focus on background scenes only and investigate whether subjective ratings of the number of details recollected differed in terms of recruitment and context conditions for only context scenes. In other words, do participants pay a high level of attention to context (in turn have higher recollected details for "Old" scenes) and how does it differ for in-person and M Turk

sample of older adults? A 4 (Context: Old vs. Novel, Novel 1) X 2 (Response Type: Old vs. New) X 2 (Experiment: In-Person vs. M Turk) mixed ANOVA with context and response as a within-subject and recruitment as a between-subject factor was conducted to investigate the effect of context conditions, age and response type during old/new judgment phase on familiarity ratings. This analysis considered whether the original, repaired, and novel context is *seen at the cued-recall test* were correctly endorsed as "Old" or incorrectly endorsed as "New."

Main Effects: There was a significant main effect of response F (1, 43) = 25.56, p <0.00 and a paired sample across responses revealed scenes with old responses (M = 2.67, SD = 1.81) had higher ratings than new responses (M = 1.27, SD = 0.79), t (44) = 5.54, p <0.00.

Interaction: No significant effects were found, context by recruitment, F (3, 129) = 0.31, p = 0.82, response type by recruitment, F (1, 43) = 0.61, p = 0.44, context by response type, F (3, 129) = 1.39, p = 0.25, context by response by recruitment, F (3, 129) = 0.08, p = 0.98.

Replicating subjective ratings of recollection for context scenes, results suggest when explicit instructions to avoid context are given, older adults regardless of sample differences, pay attention to the actuality of the instruction. Hence, while they do not explicitly pay attention to the context scenes, implicitly at least (hinted by their subjective ratings for the pairs) suggest they pay partial attention to the scenes and form instead of a vague context-pair association. Thus, their ratings for the feeling of "oldness" of context scenes are not influenced by context conditions or sample type. It might overall suggest that explicit instruction has helped older adults to avoid context scenes to form any strong subjective feeling of "oldness."

Experiment 2 Discussion

The purpose of Experiment 2 was to examine whether explicitly directing attention away from context scenes would improve FOK miscalibration for the pairs in older adults. We found explicit instruction to avoid context scenes seem to increase age-related deficiency of suppression ability, but perhaps also in a supportive manner. For FOK ratings we saw an age and context effect where older adults were well-calibrated in their predictions for recalled picture pairs across each context condition compared to younger adults. This means that despite asking them to direct away their attention, older adults pay attention to context, and perhaps more importantly, this attention to the context of all kinds seems to enhance the confidence of their prediction ratings for recalled items, especially greater for old contexts compared to new contexts. Comparing their ratings with their cued recall answers (whether they recalled the answers or not), it may suggest despite being overconfident, context might be helping older adults to calibrate accurately as their high ratings align with items they have recalled successfully.

From our objective memory of discrimination scores, we found just as Experiment 1, there was no significant effect of discrimination scores between target and lures for age groups or context. Although there is the absence of age or context effect from cued recall, we know at least that older adults pay some implicit attention to context scenes (as can be seen from their higher mean pattern for discrimination scores) due to distracted attention to context scenes when given instructions to suppress or ignore the context. This distracted attention to context scenes may help older adults FOK predictions. These age effects existed even after controlling for covariates and showed no difference in recruitment groups within older adults, which implies

that sample differences within older adults or differences in education, income, or self-esteem level have not influenced the age effects we have found between younger and older adults. Mean patterns of discrimination scores will be discussed in the General Discussion section.

Analysis with subjective ratings also supported our hypothesis that explicit instruction to direct away attention, in turn, enhances distracted attention to context scenes. We found that recollection ratings for pairs overall do not differ between younger and older adults, nor within recruitment groups of older adults. While both age groups equally ignore context for their subjective recollected details between target and lures, context influence seems to play for items that were deemed as "Old" by the participants as a whole. Overall, participants seem to have more recollected details for a pair in a repaired context compared to in the original context. Analysis with recruitment differences further shows this effect is mainly driven by older adults' in person who can accurately calibrate "old" judgments with higher recollection ratings. M Turk sample mistakenly calibrates "new" judgments with higher ratings. Although M Turk sample seems to have better recall than in-person OA, a high FA rate might be driving the later scores since cued recall is right after the study phase. While their short-term memory might be better, it seems like the M Turk sample does not pay strong attention to last the memory for later tests. Previous research has shown that reinstating context can both increase memory accuracies for targets and at the same time enhance memory illusions for lures. We think perhaps reinstating any kind of old context might have triggered both a context-pair association for older adults in-person but at the same time trigger more miscalibrated ratings for new items in M Turk sample. There could also be other reasons for miscalibration such as a lack of attention to the study from M Turk sample or the fact that M Turk sample has lower cognitive reserve factors such as

education or income which inform the lower level of cognition, may have influenced such results in M Turk sample.

For familiarity ratings, we found that older adults, in general, seem to mistakenly have a feeling of oldness for lures compared to younger adults, but these inaccuracies seem to be highest for old scenes they have seen before compared to novel scenes. Further analysis between recruitment groups showed these age-related deficiency effects are mainly driven by older adults in-person while M Turk sample is not influenced by item type or context scenes. We think that due to age-related impairment in attention to detail, older adults in person mistakenly might presume perceptually similar lures to be the same as target pairs. Thus, old context (whether original or repaired) paired with perceptually similar lures may seem more familiar to them compared to pairs of new contexts. Additionally, the effect of memory illusions due to context reinstatement as suggested earlier may also play a role in increasing familiarity for lures. Further, results imply that despite explicit instruction to draw their attention away from context, older adults in person still paid partial attention to the context to make the context-pair association, and these weak binding from the study phase seem to be strong enough to influence their later familiarity ratings. When we look within the older adult sample, we find despite inaccuracies for lure calibration, older adults in person are still calibrating their old responses with higher familiarity ratings compared to M Turk sample, just as recollection.

We also found when old context scenes are reinstated, although there is a feeling of oldness or high recollected details in older adults for scenes they think were “old”, there was no recruitment effect within older adults or any effect in younger adults. Moreover, while in general older adults may vaguely remember the background scenes due to age-deficiency attentional

mechanisms and hyper-binding, across participants this effect is absent. It is possible that the attention to context across participants to be weak. This is because our results showed as subjective ratings for recollection and familiarity for context scenes particularly were not influenced by context or age groups. While partial attention may have been given to context to form a context-pair association (suggested from subjective ratings for the pairs), high attention and in turn strong memory signal was not strong enough to form a subjective feeling of "oldness" or recollect details for the context scenes.

In sum, from Experiment 2 we know when explicit instruction to avoid context is given, older adults get hyper-aware of context and hyper-bind context to form a context-pair association. These associations are strong enough to influence their FOK ratings. Attention to context scenes makes older adults have a higher prediction for each context for items they successfully recalled, compared to younger adults. For subjective memory, we see age and context effects, and it exists even when we control for covariates. We know older adults are most susceptible to memory illusions of lures as being target due to context reinstatement when they try to recollect details about the pairs in old scenes compared to new scenes. We found participants of all ages are more susceptible to these memory illusions when try to recall vague details or feelings of "oldness" of the pairs. We also found that older adults are more prone to use familiarity for their subjective ratings compared to younger adults. Further, these associations seem also to increase accuracies of confidence for old responses for in-person older adults. Older adults in-person also seem to fall to memory illusion for lures due to context reinstatement effects for familiarity ratings. The attention to context seems weak and partial as they do not influence age or recruitment groups for scenes. But they seem to be stronger to influence FOK predictions, especially when compared to a condition when participants are not

directed away from context. Although not ideal, it seems explicit instruction to avoid context enhances attention to context, specifically predictions for future memory for in-person older adults compared to M Turk sample or younger adults.

6. GENERAL DISCUSSION

The present study investigated the role of context on FOK judgments in younger and older adults. Based on previous research that showed influence of context on unsuccessful recalled items and confidence judgments, we predicted context would also influence FOK predictions and subsequent memory stages. We found context influences FOK ratings, but unlike previous studies only in certain conditions and under certain mechanisms. Answering to our first question on how context influences FOK ratings, we found FOK judgments are influenced by old context rather than novel context (experiment 2). This is evident where participants gave higher ratings for scenes they have seen before compared to a novel context. This is also shown from our prediction analysis in Experiment 1 where we found recollection and familiarity of old scenes seem to predict FOK ratings more than new scenes. We also found an age effect to influence this FOK-context association. Specifically, we found old context influences FOK ratings more in older adults compared to younger adults for novel context conditions. Older adults compared to younger adults were particularly well calibrated in items they could recall, and which were paired with familiar scenes. Further, direct influence of age and context on FOK ratings only seem to have occurred in Experiment 2, where explicit instructions were given to avoid context.

We started off our experiment with the initial question to examine the underlying mechanism of the context-FOK relationship. Previous research on FOK judgments have shown that FOK judgments are special unlike other predictive ratings because they provide target relate information not only by direct cues (Hart 1965) but also by indirect cues and partial information of cues (Hanczakowski et al., 2014, 2015, 2017; Koriat, 1993; Schwartz & Metcalfe, 1992, 1993). The present study supports previous studies on the influence of indirect cues on FOK judgments, and we showed that familiar context as an indirect cue can also influence FOK ratings. Unlike previous studies, we showed this context influence on FOK judgments does not occur only automatically but also when participants, especially older adults seem to pay attention to context as shown by their mean discrimination scores which were higher than younger adults (more discussed in later sections). Additionally, this context influence affects older adults more than younger adults. While previous studies have shown directing attention towards context can be a supportive aid in enhancing prediction for unrecalled items and memory accuracy in older adults, in our study we found the opposite. We found instead of directing attention towards context, directing away attention provides a supportive aid to use context inform FOK predictive ratings for correctly recalled items, especially in older adults. Thus, by showing context familiarity, reinstatement, and aging effect, our current study extends the knowledge of the context-FOK relationship.

In Experiment 1, we wondered why an older adult would not heed context when previous studies have shown context influence on FOK ratings and subsequent recognition memory? Our context scenes, unlike previous studies, were not clear-cut images, to begin with, but rather a little grainy. We did not provide a clear-cut image first, because these were supposed to be the

irrelevant context that provides no target information, and second, we wanted to replicate a real-life scenario where usually context or background scenery diminishes in clarity when people focus on the task at hand. For example, if someone is having an important conversation with a friend at a coffee shop, the person would usually focus on the conversation or the friend in front of them, while the surrounding environment is pushed in the background. Both Treisman's (1964) and Broadbent's (1958) model of attention inform us that our attentional capacity is limited which makes us prioritize some attention over the others. In our study, we wanted to make sure that the task at hand and background is clear and distinguishable so that participants do not fall into the cocktail party phenomenon. We found from Experiment 1 results that such background context scenes had no influence when participants made their FOK ratings, cued recall, or recognition memory.

However, we found context did influence subjective ratings of memory, recollection, and familiarity and this differed between age groups. Unlike Broadbent's (1958) attentional model, Treisman (1964) proposed that unattended information is also processed indirectly below our awareness of our conscious level. We think that when there is no directed attention to background information that seems irrelevant as in Experiment 1, participants may not explicitly pay attention to the background, but implicitly they may still process it. This is evident by the high amount of contextual effects we found for recollection and familiarity ratings. We found in general context or age does not influence trying to recall details for either target or lures. But when participants think a pair is "old", in their subsequent rating, they seem to have high recollection for that "old" pair" especially in older adults compared to younger adults. This implies that older adults implicitly take context into account at the study phase to influence their

subjective memory, even if they are not aware of it at the moment of studying. Previous studies on aging and context have shown that older adults usually bind context to inform them of target cues as memory support and from our study, it makes sense that older adults, at least implicitly, would use more of context to recollect details about the target. Context, then seemingly provide additional support implicitly and not explicitly, to boost their feeling of details on their successful attempts.

However, context can also cause interference (Doss et al., 2018). According to the context distortion hypothesis, reinstating familiar context scenes can sometimes cause a boost in falsely recognizing a lure as a target. We think this memory illusion also occurred for our participants, not in their objective memory but at least in their subjective memory where context has the impact. Specifically, this is evident in their familiarity with the lures and much more in older adults compared to younger adults. In Experiment 1 we found that participants, in general, have higher ratings for the feeling of "oldness" of our perceptually similar lures in old context scenes compared to new context scenes, and this is higher in older adults compared to younger adults. Older adults, in general, are more prone to use automatic and less effortful gist information or familiarity to recall memory details instead of using recollection which is a more effortful conscious process (Jennings & Jacoby, 1993; Yonelinas, 2002). These automatic processes are more likely to for memory distortion or biases than effortful processing. Thus, we suggest familiarity for lures in old scenes is due to a combination of age deficiency of attention-binding mechanisms, high reliance of familiarity to inform memory cues, and increase of hyper-binding context implicitly in older adults compared to younger adults.

After finding out that context plays no role explicitly and serve as a distractor in the study phase for participants, in Experiment 2 we wanted to examine whether manipulating explicit attention away from distractors would have any influence on FOK predictions, objective memory, and implicit memory. Interestingly, we found when the context is labeled as a distractor, participants, especially older adults, explicitly pay more attention to it. This is evident in our FOK ratings where older adults had a subsequent higher rating in all three context conditions only for items they successfully recalled at cued-recall test, compared to younger adults. Perhaps at cued recall, their recall depended on remembering the context-pair association they made during the study phase. Regardless of no context or age effects at cued recall, once older adults could recall not only the target but also the correct target-context association, for their subsequent FOK prediction, their confidence increased. This is then reflected in the FOK pair ratings, greater for original context>repaired>novel context, compared to younger adults.

We also found in Experiment 2 from mean patterns that discrimination scores to be higher for older adults compared to younger adults unlike Experiment 1. Our results were surprising since previous research suggested explicit instruction *towards* context has the context influence on explicit memory. Instead of taking a stance since there is no significance, we tentatively argue that both instruction type can inform participants to use context and influence FOK predictions and memory in general. We think in our case when we asked participants to direct away attention from distractors, this instruction itself may have made it harder to guide older adults' already reduced suppression ability. This may have prompted them to use hyper-binding binding to a greater extent so that it not only influenced their implicit memory but also their explicit memory to a greater extent.

The beneficial effect of context reinstatement may also play an effect since within older adults the means suggest discriminating between target and lures were better in original context>repaired>novel context scenes. Original scenes may have brought more context-pair associative details than other scenes, particularly for older adults if they rely on context as memory aids. The age effect could suggest that although this effect of context in explicit memory in older adults is not as good as younger adults, context effects may still improve older adults' discrimination ability in general. Thus, showing that context can still influence FOK ratings and tentatively objective memory in older adults despite directing attention away from it, our current study extends the knowledge of how attentional effects of context can influence FOK judgments and memory in aging.

Like Experiment 1, for subjective memory, we found context also influenced implicit memory in older adults showing regardless of instruction manipulation the effect of context on implicit memory is a common phenomenon. Supporting previous research, we found older adults rely more on familiarity compared to younger adults and that older adults are also prone to falling into the memory illusions for context reinstatement effects. This is reflected in their higher ratings for lure pairs in old scenes compared to new scenes, suggesting age deficiencies in binding ability. We also found comparing age groups that participants, in general, have better recollection for items they think are "old" compared to "new" items, but this seems to be better for pairs in repaired context compared for pairs in the original context. Although we do not know the reasoning behind this small effect, we can say for sure that for both younger and older adults, context helps to bring more details of context-pair association to mind, more for old scenes compared to new scenes. Considering the strength of association of context influence, our results

from subjective ratings of context scenes suggest that attention to context could have been strong and partial, strong in the sense to influence FOK predictions but partial in the sense that it affected context-pair memory but not objective memory or memory for whole of context scenes exclusively.

Besides our age group differences, we should also mention our demographic variables and recruitment differences for our experiments, both of which could have influenced our results. None of the covariates influenced our age effects, except the recollection ratings for old responses. We think that these interactions resulted from chance errors due to unequal sample size in our experiment rather than an actual effect or a covariate effect because further analysis without covariate revealed no significant age effects and the effect disappeared when we controlled for covariates. Within older adults, covariates could have influenced subjective memory results due to recruitment differences. We had two different kinds of older adult samples. While some older adults were recruited in-person, we also had older adults recruited online via M Turk. The in-person sample was more educated compared to M Turk sample in both experiments and higher in income, but low self-esteem compared to M Turk sample in Experiment 2. In experiment 2, we saw the influence of covariates within older adults in recollection response ratings for pairs. While the significant interaction of response by recruitment existed, the three-way interaction of context by a response by recruitment disappeared when we controlled for covariates. This absence suggested that context did not influence response ratings between M Turk and in-person samples. From our age group, the absence of age effects also suggests that age did not affect recollection response ratings.

Previous research has shown that high education or income can serve as a proxy for high cognitive reserve factors that in turn “buffers” memory and cognition deterioration in aging (Stern, 2002). Previous research has shown that cognitive reserve factors such as education and income or personality factors such as self-esteem positively associate with higher cognition such as memory and specifically self-esteem predicts subjective memory (Opdebeeck, Martyr & Clare, 2016; Pearman & Storandt, 2004). The covariate-recruitment effects were evident in our subjective memory rating where in general we found M Turk sample having higher ratings for “new” responses or sometimes showing memory illusions to higher lures compared to older adults in-person. Although reinstatement effect of context could have played a role for the lure to old context scenes, we also think M Turk sample may have been more prone to memory illusions due to lower cognition as hinted through their covariate scores. Additionally, lower executive function attention to details may have also made them have poor calibration skills between new responses in old/new judgment and subsequent perceived ratings for those pairs as shown by the high recollection responses for old items in context conditions. This is also evident when the 3-way interaction disappears as we control for covariates suggesting the context effects occur only due to differences in covariate factors between the two-sample type. We cannot directly make assumptions for self-esteem scores as the previous study only showed a correlation effect, not causation and in further suggested, depression could be a mediating factor between self-esteem and memory. In our study, none of the older adults included were depressed (as shown via the GDS scale, see Table 2) nor had any influence on the effects of Experiment 1. Although high self-esteem may hint at better memory it can also cause vulnerability to more memory errors (Szpitalak & Polczyk, 2013). We think in our case within older adults, high self-esteem may have had the latter effect on M Turk sample as high self may have made M Turk sample

overconfident (evident in their high FOK ratings compared to the in-person sample in Experiment 1) which could have led to more errors compared to older adults in-person. This may also suggest that in Experiment 1 while the absence of age effects could be generalizable, the absence of response effect might not be generalizable. Thus, future studies should recruit a sample of older adults with a uniform demographic to test for response differences in FOK ratings.

Implications and Future Directions

From two experiments we have found that older adults take context into account while making their memory judgment under certain conditions. We found older adults only take explicit attention to context for their FOK judgments when we ask them not to pay attention to it. Results overall suggest both types of contextual instruction manipulation to direct attention, towards context and away from it, can influence FOK predictions in general. Since the current study had recruitment differences, future studies should replicate the experiment with a uniform sample to test for generalizability. Future research should also consider looking at FOK accuracies, aka gamma scores. As a reminder, in the current study we are not interested in participants' FOK accuracy but rather what do they consider when making their FOK ratings. So, while FOK accuracy is not relevant to the current study, we think with FOK accuracy, we can dive deeper to understand exactly what item participants pay attention to between cue and target when the context is reinstated for FOK judgments. Is it the association with the cue-context or target-context? If context reinstatement increases FOK mean ratings without affecting their accuracy, then it can suggest FOK judgments depend on retrieval of information about the cue-context, as opposed to the target-context.

Future research should also consider a sample with more diversity to consider how would culture influences how people make memory predictions and subsequent memory with context. For example, research shows Eastern cultures show a more holistic approach and pay more attention to context compared to Western cultures who show a more rule-based/categorical approach. But this difference in approach can also depend on resources and task demands with age (Park et al., 1999). In younger adults, we may see cross-cultural differences, but they may converge in the aging population due to decreased cognitive flexibility across the lifespan. Further, it can also depend on automatic vs. effortful processing. East Asian young may already be relying on automatic processes for attending to detail and thus, the decline in East Asians with age (for example in recollection in our case) would be less than for Westerners who are relying on more effortful processes at both ages. Thus, for our study, previous research may hint an age decline of total subjective memory in East Asians compared to Westerners.

Understanding how FOK is influenced by context is important as we age. FOK ratings are important because this is where we see the most miscalibration between old and young. We need to address how we can improve it because when the tip-of-the-tongue phenomenon increases with aging, it can lead to an increase in frustrations, anger, reduced confidence in oneself, low self-esteem, which in turn can reduce the performance of the remaining intact memory. As background cues have been shown to act as a memory aid in aging, we must use such helpful cues to understand how we can not only improve older adults' memory but also their perception of their memory, which is the underlying basis for good memory accuracies.

7. REFERENCES

Anseel, F., Lievens, F., & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, *110*(1), 23-35.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640-1642.

Bäckman, L., & Dixon, R. A. (1992). Psychological compensation: a theoretical framework. *Psychological bulletin*, *112*(2), 259.

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience*, *28*(34), 8539-8544.

Biss, R. K., Ngo, K. J., Hasher, L., Campbell, K. L., & Rowe, G. (2013). Distraction can reduce age-related forgetting. *Psychological Science*, *24*(4), 448-455.

Burke, S. N., Gaynor, L. S., Barnes, C. A., Bauer, R. M., Bizon, J. L., Roberson, E. D., & Ryan, L. (2018). Shared functions of perirhinal and parahippocampal cortices: implications for cognitive aging. *Trends in neurosciences*, *41*(6), 349-359.

Butler, R. N. (2010). *The longevity prescription: The 8 proven keys to a long, healthy life*. Penguin.

Cacioppo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of personality and social psychology*, *51*(5), 1032.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of personality assessment*, *48*(3), 306-307.

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: A unique age effect. *Psychological Science*, *21*(3), 399-405.

Campbell, K. L., Grady, C. L., Ng, C., & Hasher, L. (2012). Age differences in the frontoparietal cognitive control network: implications for distractibility. *Neuropsychologia*, *50*(9), 2212-2223.

Chalfonte, B.L., Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & cognition*, *24*(4), 403-416.

Craik, F. I., & Schloerscheidt, A. M. (2011). Age-related differences in recognition memory: Effects of materials and context change. *Psychology and Aging*, *26*(3), 671.

Doss, M. K., Picart, J. K., & Gallo, D. A. (2018). The Dark Side of Context: Context Reinstatement Can Distort Memory. *Psychological science*, 0956797617749534.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & cognition*, *25*(5), 691-700.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Haugtvedt, C. P., Petty, R. E., & Cacioppo, J. T. (1992). Need for cognition and advertising: Understanding the role of personality variables in consumer behavior. *Journal of Consumer Psychology*, *1*(3), 239-260.

Howland, M., Tatsuoka, C., Smyth, K. A., & Sajatovic, M. (2016). Detecting Change over Time: A Comparison of the SLUMS Exam and the MMSE in Older Adults At Risk for Cognitive Decline: Stability of the SLUMS Exam versus the MMSE. *CNS Neuroscience & Therapeutics*, *22*(5), 413–419. http://doi.org/10.1111/cns.12515

Holden, H. M., Hoebel, C., Loftis, K., & Gilbert, P. E. (2012). Spatial pattern separation in cognitively normal young and older adults. *Hippocampus*, *22*(9), 1826-1832.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language*, *58*(4), 1095-1111.

Garrett, D. D., Grady, C. L., & Hasher, L. (2010). Everyday memory compensation: the impact of cognitive reserve, subjective memory, and stress. *Psychology and Aging*, *25*(1), 74.

Gopie, N., Craik, F. I., & Hasher, L. (2011). A double dissociation of implicit and explicit memory in younger and older adults. *Psychological Science*, *22*(5), 634-640.

Gutchess, A. H., & Park, D. C. (2009). Effects of ageing on associative memory for related and unrelated pictures. *European Journal of Cognitive Psychology*, *21*(2-3), 235-254.

Hanczakowski, M., Zawadzka, K., & Cockcroft-McKay, C. (2014). Feeling of knowing and restudy choices. *Psychonomic bulletin & review*, *21*(6), 1617-1622.

Hanczakowski, M., Zawadzka, K., Collie, H., & Macken, B. (2017). Metamemory in a familiar place: The effects of environmental context on feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 59.

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language*, *52*(4), 493-504.

Kirwan CB, Stark CE. 2007. Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. Learn Mem 14:625–633.

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

Koutstaal, W. (2003). Older adults encode—but do not always use—perceptual details:

Intentional versus unintentional effects of detail on memory judgments. *Psychological Science*, *14*(2), 189-193.

Luo, L., Hendriks, T., & Craik, F. I. (2007). Age differences in recollection: Three patterns of enhanced encoding. *Psychology and Aging*, *22*(2), 269.

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 153.

Memel, M., & Ryan, L. (2017). Visual integration enhances associative memory equally for young and older adults without reducing hippocampal encoding activation. *Neuropsychologia*, *100*, 195-206.

McDonough, I. M., & Gallo, D. A. (2008). Autobiographical elaboration reduces memory distortion: Cognitive operations and the distinctiveness heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1430.

Medina, L., Sabo, S., & Vespa, J. Living Longer: Historical and Projected Life Expectancy in the United States, 1960 to 2060.

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.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519-533.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170.

Naveh-Benjamin M, Hussain Z, Guez J, Bar-On M (2003). Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 826–837.

Naveh-Benjamin, M., Guez, J., Kilb, A., & Reedy, S. (2004). The associative memory deficit of older adults: Further support using face-name associations. *Psychology and aging*, *19*(3), 541.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation*, *26*, 125-141.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*(6), 838-848.

Opdebeeck, C., Martyr, A., & Clare, L. (2016). Cognitive reserve and cognitive function in healthy older people: a meta-analysis. *Aging, Neuropsychology, and Cognition*, *23*(1),

40-60.

Park, D. C., Smith, A. D., Morrell, R. W., Puglisi, J. T., & Dudley, W. N. (1990). Effects of contextual integration on recall of pictures by older adults. *Journal of gerontology*, *45*(2), 52-57.

Park, D. C., Puglisi, J. T., & Sovacool, M. (1984). Picture memory in older adults: Effects of contextual detail at encoding and retrieval. *Journal of Gerontology*, *39*(2), 213-215.

Pearman, A., & Storandt, M. (2004). Predictors of subjective memory in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *59*(1), 4-6.

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.

Reinhard, M. A., & Dickhäuser, O. (2009). Need for cognition, task difficulty, and the formation of performance expectancies. *Journal of Personality and Social Psychology*, *96*(5), 1062.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive psychology*, *19*(1), 90-138.

Reder, L. M., Victoria, L. W., Manelis, A., Oates, J. M., Dutcher, J. M., Bates, J. T., ... & Gyulai, F. (2013). Why it's easier to remember seeing a face we already know than one we don't: Preexisting memory representations facilitate memory formation. *Psychological Science*, *24*(3), 363-372.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2017). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability.

Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy. Measures package*, *61*, 52.

Rowe, G., Valderrama, S., Hasher, L., & Lenartowicz, A. (2006). Attentional disregulation: A benefit for implicit memory. *Psychology and aging*, *21*(4), 826.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1074.

Singer, M., & Tiede, H. L. (2008). Feeling of knowing and duration of unsuccessful memory search. *Memory & Cognition*, *36*(3), 588-597.

Smith, S. M., & Manzano, I. (2010). Video context-dependent recall. *Behavior Research Methods*, *42*(1), 292-301.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, *8*(2), 203-220.

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

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.

Szpitalak, M., & Polczyk, R. (2013). Promoting eyewitness testimony quality: Warning vs. reinforced self-affirmation as methods of reduction of the misinformation effect. *Polish Psychological Bulletin*, *44*(1), 85-91.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the international neuropsychological society*, *8*(3), 448-460.

Tariq, S. H., Tumosa, N., Chibnall, J. T., Perry III, M. H., & Morley, J. E. (2006). Comparison of the Saint Louis University mental status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder—a pilot study. *The American journal of geriatric psychiatry*, *14*(11), 900-910.

Thomas, R. C., & Hasher, L. (2012). Reflections of distraction in memory: Transfer of previous distraction improves recall in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 30.

Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2009). Visual object pattern separation deficits in nondemented older adults. *Learning & memory*, *16*(5), 338-342.

Wechsler, D., Coalson, D. L., & Raiford, S. E. (1997). *WAIS-III: Wechsler adult intelligence scale*. San Antonio, TX: Psychological Corporation.

Weeks, J. C., Biss, R. K., Murphy, K. J., & Hasher, L. (2016). Face–name learning in older adults: A benefit of hyper-binding. *Psychonomic Bulletin & Review*, *23*(5), 1559-1565.

Weeks, J. C., & Hasher, L. (2014). The disruptive–and beneficial–effects of distraction on older adults' cognitive performance. *Frontiers in psychology*, *5*, 133.

Wilkinson, G. S. (1993). *WRAT-3: Wide range achievement test administration manual*. Wide Range, Incorporated.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in neurosciences*, *34*(10), 515-525.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of psychiatric research*, *17*(1), 37-49.

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.

8. APPENDIX A

Table 2

Demographics Experiment 1 and 2

	Experiment 1				Experiment 2			
	Young	*Old All*	*Old-in person*	*Old-M Turk*	*Young*	*Old All*	*Old-in person*	*Old-M Turk*
	N	*N*	*N*	*N*	*N*	*N*	*N*	*N*
N=	30.00	59.00	30.00	29.00	30.00	45.00	16.00	29.00
Gender								
Male	4.00	20.00	7.00	13.00	5.00	20.00	5.00	15.00
Female	26.00	39.00	23.00	16.00	25.00	25.00	11.00	14.00
Race								
Caucasian	29.00	52.00	27.00	25.00	26.00	44.00	15.00	29.00
African American	1.00	5.00	2.00	3.00	1.00	1.00	1.00	0.00
Asian	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hispanic	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other/unknown	0.00	2.00	1.00	1.00	3.00	0.00	0.00	0.00
Mean Age (in years)	18.80	67.01	70.33	63.69	18.74	65.66	67.25	64.06
SD	1.27	6.89	6.78	6.99	0.94	5.69	7.00	4.38
Mean Education (in years)	14.93	17.12	17.90	16.34	14.17	14.66	18.00	11.31
SD	1.80	2.44	2.19	2.69	2.03	4.68	3.24	6.12
Father's Education (in years)	15.00	11.06	10.53	11.59	12.80	12.48	12.13	12.82
SD	6.62	5.28	5.16	5.40	6.54	3.89	3.75	4.02
Mother's Education (in years)	10.3		12.27	11.31	12.70	11.68	10.12	13.24
SD	8.17	4.09	3.18	4.99	6.79	4.38	5.53	3.22
Average Income ($ in thousand)	8.80	5.35	6.00	4.69	7.90	4.72	5.44	4.00
SD	3.29	2.63	2.62	2.63	3.76	2.28	2.13	2.42
Verbal Score (ETS)	9.13	20.38	18.54	22.22	9.96	22.57	22.41	22.72
SD	6.88	7.08	7.14	7.02	5.01	7.01	6.69	7.33

Self Esteem Score (Rosenberg Scale)	24.17	24.57	25.00	24.14	25.00	19.22	13.44	25.00
SD	4.56	2.32	0.00	4.64	0.00	2.97	5.94	0.00
Total Reading Score (Wide Range Achievement, WRAT-4 Scale)	N/A	N/A	N/A	N/A	N/A	N/A	70	N/A
SD	N/A	N/A	N/A	N/A	N/A	N/A	0.00	N/A
Motivation (Need for Cognition Scale)	52.47	53.76	52.87	54.65	15.50	53.24	54.37	52.10
SD	10.44	12.84	10.60	15.08	7.61	10.99	9.45	12.53
Depression (Geriatric Depression Scale)	N/A	2.20	2.06	2.34	2.27	1.46	1.50	1.41
SD	N/A	1.93	1.53	2.33	1.50	1.02	0.63	1.40
Working Memory and Cognitive Decline (The Saint Louis University Mental Status Examination Scale)	N/A	N/A	29.33	N/A	N/A	N/A	29.69	N/A
SD	N/A	N/A	1.18	N/A	N/A	N/A	1.01	N/A
Working Memory and Cognitive Decline for M-Turk (Short Portable Mental Status Questionnaire)	N/A	N/A	N/A	8.00	N/A	N/A	N/A	8.00
SD	N/A	N/A	N/A	0.00	N/A	N/A	N/A	0.00

9. APPENDIX B

SPMSQ Questionnaire for M Turk

SPMSQ Screener for M Turk

1. What are the month date and year for today? (mm/dd/yyyy): _____
2. What is the day of the week? _____
3. Do you know your phone number? _____
4. Were you born before 1964? _____
5. Who is the current president? _____
6. Who was the president before him? _____
7. Do you know your mother's maiden name? _____
8. Can you count backward from 20s by 3's? Yes/No

10. APPENDIX C

IRB Approval Letter

THE UNIVERSITY OF ALABAMA | Office of the Vice President for Research & Economic Development
Office for Research Compliance

December 19, 2019

Tasnuva Enam, MA
Department of Psychology
College of Arts & Sciences
The University of Alabama
Box 870348

Re: IRB # 18-OR-246-R1-C "Influence of Age and Context on Feeling of Knowing (FOK) Judgments and Recognition Memory"

Dear Ms. Enam:

The University of Alabama Institutional Review Board has reviewed the revision to your previously approved expedited protocol. The board has approved the change in your protocol.

Please remember that your protocol will expire on August 21, 2020.

Should you need to submit any further correspondence regarding this proposal, please include the assigned IRB application number. Changes in this study cannot be initiated without IRB approval, except when necessary to eliminate apparent immediate hazards to participants.

Good luck with your research.

[redacted]

Jeniup Building | Box 870127 | Tuscaloosa, AL 35487-0127
205-348-8461 | Fax 205-348-7189 | Toll Free 1-877-820-3066