Aging and recognition memory: A meta-analysis

Scott H. Fraundorf

University of Pittsburgh

Kathleen L. Hourihan

Memorial University of Newfoundland

Rachel A. Peters

University of Pittsburgh

Aaron S. Benjamin

University of Illinois at Urbana-Champaign

Author Note

Scott H. Fraundorf, Learning Research and Development Center and Department of Psychology, University of Pittsburgh; Kathleen L. Hourihan, Department of Psychology, Memorial University of Newfoundland; Rachel A. Peters, Learning Research and Development Center, University of Pittsburgh; Aaron S. Benjamin, Department of Psychology, University of Illinois at Urbana-Champaign.

Correspondence concerning this article should be addressed to Scott Fraundorf, Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA, 15260. Phone: (412) 624-7029. Email: scottfraundorf@gmail.com

Abstract

Recognizing a stimulus as previously encountered is a crucial everyday life skill and a critical task motivating theoretical development in models of human memory. Although there are clear age-related memory deficits in tasks requiring recall or memory for context, the existence and nature of age differences in recognition memory remain unclear. The nature of any such deficits are critical to understanding the effects of age on memory because recognition tasks allow fewer strategic backdoors to supporting memory than do tasks of recall. Consequently, recognition may provide the purest measure of age-related memory deficit of all standard memory tasks. We conducted a meta-analysis of 232 prior experiments on age differences in recognition memory. As an organizing framework, we used signal-detection theory (Green & Swets, 1966; Macmillan & Creelman, 2004) to characterize recognition memory in terms of both discrimination between studied items and unstudied lures (d') and response bias or criterion (c). Relative to young adults, older adults showed reduced discrimination accuracy and a more liberal response criterion (i.e., greater tendency to term items *new*). Both of these effects were influenced by multiple, differing variables, with larger age deficits when studied material must be discriminated from familiar or related material, but smaller when studying semantically rich materials. These results support a view in which neither the self-initiation of mnemonic processes nor the deployment of strategic processes is the only source of age-related memory deficits, and they add to our understanding of the mechanisms underlying those changes.

Keywords: recognition memory; memory aging; cognitive aging; meta-analysis

Public Significance Statement: This meta-analysis indicates that older adults are less effective than young adults at recognizing previously presented stimuli, such as words, faces, or pictures. Older adults also have a stronger tendency to err on the side of judging things as old (previously studied or encountered). Differences in age deficits across tasks supports theories that older adults rely more on meaning in memory and have less accurate memory for the source of information.

Difficulties with memory are one of the major complaints of older adults (Hertzog, 2002; Lineweaver & Hertzog, 1998). Even in healthy older adults who do not suffer from organic brain disease that leads to known cognitive deficits, memory problems are common (Hoyer & Verhaeghen, 2006). A major task for cognitive psychologists, neuropsychologists, and neuroscientists who study cognitive aging is to identify patterns of impairment, to develop theories that accommodate and make sense of those patterns, and to use those theories as building blocks to develop regimes to remediate age-related memory impairment.

One established regularity of memory aging is that older adults show clear deficits relative to young adults in tasks of *free recall* (Craik & McDowd, 1987; Danckert & Craik, 2013), which require the learner to bring sophisticated encoding and retrieval strategies to bear in order to attain high levels of performance (Benjamin, 2008; Fiechter, Benjamin, & Unsworth, 2016). Advancing age also has strong deleterious effects on the ability to remember the source or context of information (Jennings & Jacoby, 1993; Old & Naveh-Benjamin, 2008; Spencer & Raz, 1995).

However, it is less clear whether there are age differences in *recognition memory*—that is, tasks in which the goal is to judge whether things have been encountered before or not, such as whether a face is a known associate or a stranger, whether a book or article has been read before, or whether a factual statement matches what has been heard before (Craik & McDowd, 1987; Danckert & Craik, 2013; Hoyer & Verhaeghen, 2006; Schonfield & Robertson, 1966; Verhaeghen, Marcoen, & Goossens, 1993). Tasks involving recognition memory are standard-bearers for uses of memory in which successful performance is thought to be guided more by the task environment, including the stimulus. As such, they provide a fascinating test case for the magnitude of deficits under conditions in which self-directed processing is minimized.

In addition, theories of recognition memory are among the most advanced and successful theories in memory research, and recognition memory has been particularly informative in advancing

theories of memory in young adults (e.g., Benjamin, 2010; Hintzman, 1988; Jacoby, 1991; Wixted, 2007; Yonelinas, 1994). Examining memory aging through the lens of recognition memory is thus theoretically direct because those theories enable us to make more precise statements about the origins of age-related deficits, if they exist. Recognition memory is a particularly useful case because even common recognition tasks permit a separation of the actual fidelity of older adults' memory from the strategic decisions that surround making mnemonic judgments. While the bulk of theorizing regarding memory aging has concerned the fidelity of memory itself, it is plausible that age could also affect the decision processes surrounding memory judgments (e.g., Benjamin, Diaz, & Wee, 2009; Cassidy & Gutchess, 2015; Criss, Aue, & Kilic, 2014; Gordon & Clark, 1974; Huh, Kramer, Gazzaley, & Delis, 2006; Kapucu, Rotello, Ready, & Seidl, 2008).

Indeed, despite decades of research into human memory, the causes and mechanisms of memory aging—both in recognition memory and more broadly—remain debated. For example, theories differ on whether age differences might stem from a single global deficit in memory (e.g., Benjamin, 2010, 2016; Salthouse, 1996) or from the effects of some processes, such as associative memory, differentially declining more than others (Jennings & Jacoby, 1993; Old & Naveh-Benjamin, 2008; Smyth & Naveh-Benjamin, 2016; Zacks & Hasher, 2006). Further, it is unclear whether a single age-related mechanism is sufficient to account for all age-related differences in memory or if the effects of age on memory reflect the combination of multiple age-related changes—such as both increases in semantic knowledge and a motivational shift towards prioritizing positive affect—acting in concert (e.g., Healey & Kahana, 2016).

In this article, we provide the first-ever meta-analysis of age-related effects on general recognition memory, with special attention paid to issues that have not been deeply explored in past work. We examine 232 studies of age differences in recognition memory to determine whether there is a general deficit in recognition memory in older adults as compared to younger adults. In addition, we

test, across studies, many potential moderators of these age deficits that have been suggested by the literature; these moderators are relevant to evaluating competing theoretical claims about the origin of age-related differences in episodic memory and consequently about which variables should moderate these changes. This work is relevant both to the applied problem of characterizing the problems of memory that accompany aging and to the theoretical problem of developing theories of recognition, and theories of the effects of aging upon recognition. But, we first pause to lay out our theoretical framework for assessing and understanding recognition, and also to explain how this current meta-analysis fits into the larger body of research on aging and memory.

Discrimination and Response Criteria in Recognition Memory

Recognition memory poses a particularly informative source of evidence for theories of memory aging because it exposes the strategic decisions in how people approach memory tasks. In judging whether items have been previously encountered, rememberers are guided not just by the fidelity of their memory but also by their preference to make certain types of responses or avoid certain types of errors. For instance, imagine an older adult who passes a person on the street who looks somewhat familiar and consequently waves hello. The decision to wave is likely influenced, in part, by how confident that our subject is that the passerby is indeed an acquaintance. However, a second reason for waving may be that our subject is worried about declines in his or her memory and thus, to avoid an embarrassing instance of forgetting, prefers to wave when uncertain even if it means occasionally waving to a stranger. By contrast, a younger individual might err on the side of *not* waving to avoid an awkward incident of waving to a stranger, even if it means occasionally failing to greet an acquaintance.

This intuition is captured by detection-theoretic accounts of recognition memory in which recognition memory judgments can be decomposed at least two components. (For a general account of signal-detection theory, see Green & Swets, 1966; Macmillan & Creelman, 2004; for applications to

recognition memory in particular, see Banks, 1970; Egan, 1958; Lockhart & Murdock, 1970; Parks, 1966.) *Discrimination* or *sensitivity* captures how well people can successfully discriminate old (studied) items from new (unstudied) items; that is, it measures how much more likely it is that people will judge an item as *old* when the item is actually old as compared to when it is new. By contrast, *response criterion* or *response bias* captures the criterial level of evidence at which a learner terms an item *old*; that is, it measures whether there is any overall preference to term items *old* or *new* independent of their studied status.

   In our meta-analysis, we index these properties with the measures *d'* and *c*. A larger *d'* indicates superior discrimination; a larger *c* indicates a more conservative criterion in which learners err on the side of judging items as new and unstudied whereas a smaller *c* indicates a liberal criterion in which learners err on the side of judging items as old and unstudied. These measures likely simplify the true cognitive processes involved in recognition memory: For instance, many studies have suggested that the distribution of strength for old items actually appears to have larger variance than the distribution of strength for new items (Ratcliff, Sheu, & Gronlund, 1992). This finding has been interpreted to indicate greater variability in the encoding process (Wixted, 2007) or the contribution of an all-or-none recollective process that operates for some but not all items (Yonelinas, 1994). Other theories extend the signal-detection model with additional parameters, such as the degree to which the recognition criterion shifts across trials (Benjamin, Diaz, & Wee, 2009). Unfortunately, the type of data that would allow estimation of any one of these more complex models appears far too infrequently in the literature to permit a meta-analysis of age differences. Thus, we thus use *d'* and *c* because they represent parameters from the best possible validated and interpretable model of decision-making in recognition given the available data. Nonetheless, it should be kept in mind that they are likely simplified representations of the exact criterion placement or discriminability for an individual or condition, and they are best interpreted descriptive indices that allow us to compare general

discrimination ability and response bias across a variety of study and test conditions.

Past work has largely focused (explicitly or implicitly) on age differences in discrimination. However, it is plausible that criterion could also shift with age—for instance, the stereotype that age is associated with forgetting might motivate a shift towards affirming items as previously encountered, so as to avoid "forgetting." Indeed, it has been suggested that some effects previously viewed as age differences in discrimination may in fact have stemmed from age differences in criterion placement that were not accounted for (e.g., Kapucu et al., 2008). Thus, controlling for and examining potential age differences in criterion placement is necessary to provide a full account of memory aging. In addition, age differences in criterion (or lack thereof) are informative for evaluating different theories of memory aging, which in some cases make competing predictions about whether there should be age differences in criterion placement and, if so, what might influence them. We discuss those theories, and their broader predictions, below.

Theories of Age Differences in Episodic Memory

Many different theories have been developed in the literature to account for age differences in episodic memory; these theories make conflicting claims in some cases but overlap in others. We discuss here six classes of theories that can be assessed with the present data; in the Discussion, we also briefly note additional prominent theories that fall outside the scope of the meta-analysis. Because the vast majority of theories in this field have been stated only verbally, it can be difficult to pin down specific predictions and points of contrast. We have tried to take as broad and dispassionate a view of those perspectives as possible, but some may certainly disagree with our interpretations and consequently disagree with the implications of the meta-analysis for those theories. We hope that this work will also help sharpen the field's understanding of where our theories are underdeveloped and spur additional work on those soft points in our understanding.

One prominent class of theories, which we term *process* theories, attribute age-related

differences in memory to the impairment of specific cognitive processes. For instance, one prominent

theory (Yonelinas, 1994) posits two processes underlying recognition memory judgments: a feeling of

general *familiarity* that does not make reference to a particular study episode and a *recollection* process

that actively reconstructs a particular prior episode. Recollection is hypothesized to decline especially

strongly with age whereas familiarity is thought to be spared (Jennings & Jacoby, 1993). This theory

accounts for the general observation, supported by past meta-analyses, that older adults have particular

difficulty with tasks that are partially or wholly dependent upon remembering the source or context of

information (Spencer & Raz, 1995; Old & Naveh-Benjamin, 2008). For instance, some experiments

task participants with rejecting *conjunction lures* made up of two previously studied parts, such as two

studied word pairs reassembled (e.g., *SNOW—GOLF* and *PROJECTOR—STONE* reassembled to form

SNOW—STONE). Because all of the components of this lure have been seen before, they cannot be

rejected on the basis of familiarity, and recollection is required. Thus, this theory predicts that older

adults should exhibit especially poor memory for tasks that require rejecting conjunction lures.

A related view is that older adults are particularly poor at *initiating* mnemonic processing on

their own (e.g., Craik, 1986; Luo & Craik, 2008). Thus, older adults are especially disadvantaged

relative to young adults when a task requires them to initiate their own encoding or retrieval strategies,

but age differences are relatively small when the environment contains valid cues that can support

performance. This self-initiated processing theory predicts larger age differences, for instance, when

learners must self-regulate their study by choosing how long to study each item. Conversely, age

differences should be relatively small when the task guides learners to the use of efficacious strategies,

such as deep, elaborative encoding; such tasks minimize the self-initiation required and assure that all

learners, young and old alike, are using the best strategies. (Indeed, guiding or requiring elaborative

encoding has been proposed as a strategy to ameliorate age differences; Craik, 1986).

A third process that has been claimed to be particularly degraded by memory aging is inhibition

(Hasher, Zacks, & May, 1999; Healey, Campbell, & Hasher, 2008). Older adults may be less proficient at removing outdated information from the focus of attention and at preventing irrelevant information from being attended in the first place. Although perhaps most directly applicable to intrusions in recall and to working memory tasks that require sustained attention, deficits in inhibition could also have consequences for long-term recognition memory. For instance, reduced inhibition of memory responses may lead older adults to erroneously affirm too many unstudied lure items, thus resulting in a more liberal response criterion (Huh et al., 2006).

By contrast, what we term *global deficit* accounts claim that there are no specific mnemonic processes that are particularly spared or impaired with age. Rather, memory fidelity is globally decreased, perhaps because of more general cognitive changes, such as a decline in basic processing speed that impairs memory by delaying transmissions within or across neural systems (Salthouse, 1996). These theories predict that recognition memory discrimination should be generally reduced in older adults as compared to younger adults, for all tasks and stimuli. In addition, global deficit accounts can explain declines in remembering particular kinds of information, such as the age-related deficit in contextual memory discussed above, if that information is assumed to be represented less redundantly. For instance, in the DRYAD model (Benjamin, 2010, 2016; Benjamin, Diaz, Matzen, & Johnson, 2012), central "item" information is assumed to be encoded with some degree of redundancy, so losing part of the representation does not greatly impair those memories. Contextual information, by virtue of being less central to the material being studied, is encoded less redundantly. Consequently, any source of noise in the representation—such as general cognitive declines with age—impairs contextual memory more than item memory, even if no separate neural or cognitive *process* is posited for remembering contexts or sources. Models such as DRYAD thus imply that age differences in recognition memory should be particularly stark for difficult tasks (i.e., those in which discrimination is generally poor) because the global age deficit will be particularly destructive for poor-quality, less

redundant memory representations.

Another broad claim about older adults' memory that has been widely advanced is that it is more reliant upon semantic knowledge than younger adults' memory (Zacks & Hasher, 2006), perhaps because older adults have simply acquired more such knowledge over the course of their lives. This emphasis on semantic representations can support veridical memory when the memoranda are tied into existing semantic or world knowledge (e.g., Castel, 2005; Castel, McGillivray, & Worden, 2013; McGillivray & Castel, 2017), but can also lead to confusions among items that are semantically similar. Thus, semantic-processing theories predict that older adults' ability to discriminate previously studied information should be relatively spared for semantically richer materials (e.g., pictures or extended texts as compared to isolated words) but should be particularly poor in recognition tasks that require discriminating studied targets from unstudied lures that are nevertheless related to the target at a semantic or meaning level (e.g., two synonymous words).

Another class of theories, which we term *motivational*, is that age differences in recognition memory performance are driven not by older adults' *ability* to perform such tasks per se, but rather that older adults are not motivated by the types of stimuli and tasks used in typical laboratory studies. For instance, Castel and colleagues (Castel, 2007; Castel, Benjamin, Craik, & Watkins, 2002; Castel, Farb, & Craik, 2007) have argued that older adults are superior at identifying and focusing on high-value material; thus, older adults might show poor memory for laboratory stimuli because they have identified them as irrelevant or uninteresting. That is, if memory tasks were more relevant to older adults' lives, they would be more motivated to perform well, and might perform similarly or equivalent to young adults. In particular, socioemotional selectivity theory (Carstensen, Isaacowitz, & Charles, 1999) proposes that older adults perceive their time as more limited than younger adults and consequently are motivated to spend their time on experiencing positive emotions and on reviewing known information rather than on acquiring new information (Mather, 2004). This theory predicts that

older adults' memory should be comparatively good (relative to young adults) for positive emotional stimuli and should be comparatively poor for negatively-valenced stimuli.

A final claim is that older adults' performance in laboratory memory tasks is affected by the phenomenon of *stereotype threat* (Steele & Aronson, 1995), in which activation of a personally relevant negative stereotype results in declines in cognitive performance. For instance, a general cultural stereotype is that memory declines with age and that older adults frequently forget things (Hertzog, 2002; Lineweaver & Hertzog, 1998). This stereotype is then activated by the researchers' recruitment materials, their task instructions, or even the laboratory name, disadvantaging older adults' task performance relative to young adults'. This claim has been supported by evidence that age differences in memory can be reduced or eliminated by eliminating references that might activate the stereotype, such as by framing the experimental task as involving "learning" rather than "memory" (Rahhal, Hasher, & Colcombe, 2001) and by evidence that age differences are *enhanced* by explicitly presenting the stereotypes within the experimental procedure (Barber & Mather, 2013). We will have relatively little to say about the stereotype-threat account in our meta-analysis because relatively few researchers manipulate or report sufficient details of their instructions. However, we note that this account suggests that older adults who are particularly worried about forgetting might have a comparatively liberal response criterion in which they err on the side of terming items *old* or *studied* so as not to inadvertently miss an old/studied item and thus confirm the stereotype.

The Meta-Analytic Approach

The theories described above have often been tested in separate studies. For instance, to evaluate self-initiated processing accounts, researchers might manipulate whether a deep, elaborative processing task is required and test whether this manipulation differentially affects older adults relative to young adults; meanwhile, affective valence is either held constant or not controlled at all. These experiments have contributed much to our understanding of the aging of memory. But one challenge is

that testing individual variables and hypotheses in isolation makes it difficult to assess potential influences in conjunction with each other and whether one or more than one mechanism is necessary to fully account for age differences in memory. Further, individual experiments in cognitive psychology often have only moderate power (Open Science Framework, 2015), which means that the outcomes of many individual experiments are likely to be spurious (Ioannidis, 2005).

Alternately, various forms of meta-analysis synthesize existing data from multiple experiments in a single statistical analysis. Meta-analyses have many important advantages (e.g., Cooper & Hedges, 1994): By pooling data from many studies, they provide a more powerful assessment of the robustness of a putative effect across experiments, and they are able to obtain more precise estimates of effect size. The larger sample sizes characteristic of a meta-analysis also permit testing whether a given effect (e.g., age differences in recognition memory discrimination) is moderated by other variables, such as the emotional valence of the stimuli. Further, meta-analyses can even test hypotheses not necessarily present in the original papers, such as comparing whether age differences vary in magnitude across experiments involving memory for faces versus memory for pictures versus memory for words. Finally, relative to qualitative reviews of the literature, meta-analyses have the advantage that their conclusions are supported by quantitative information.

Meta-analyses are not without their own challenges. Analyses of existing data are necessarily limited to the subject populations and measures contained in those datasets, and they do not permit direct experimental manipulations that justify claims of causality. Further, any publication bias that exists in the literature (e.g., that positive findings are more likely to be published than negative findings) would lead to a non-representative sample of all relevant studies. In our analysis, we assess evidence for or against publication bias in this literature. And, in the Discussion, we revisit these limitations, their application to the present study, and how they do or do not temper the conclusions.

Present Study

To better characterize age differences in recognition memory and their potential sources, we conducted a meta-analysis of past experiments with cross-sectional comparisons of young and healthy older adults engaging in tasks of recognition memory. Past meta-analyses of age differences in memory have either examined memory processes more broadly and could not take a close look at mediators of recognition memory (Verhaeghen et al., 1993), or they have more narrowly investigated age differences specifically in tasks that require remembering the context or source of information (Spencer & Raz, 1995; Old & Naveh-Benjamin, 2008) or that involve memory for faces of different ages (Rhodes & Anastasi, 2011). Here, we target an intermediate level of analysis by examining specifically recognition memory but examining the broad spectrum of recognition memory tasks, which allowed us to characterize potential mediators of any age differences. We focus on cross-sectional rather than longitudinal comparisons because they are more numerous in the literature, especially for the sort of parametric experimental manipulations (e.g., of deep versus shallow encoding) examined here; in the Discussion section, we describe some of the strengths and weaknesses of cross-sectional comparisons.

We examine, first, whether there are overall age differences in recognition memory discrimination and response criterion placement[1], as characterized by d' and c. We then further examine whether these age differences are enhanced or reduced by many of the variables commonly varied in memory studies, such as the type of memoranda (e.g., faces versus words versus pictures), the affective valence of their stimuli, the retention interval, and the types of lures from which the studied information must be discriminated. (See the Method section, below, for a detailed description of all the relevant variables or Table 2 for a summary.) In selecting these variables, we are guided both by the specific theories of memory aging reviewed above and by the general set of features that commonly

---

[1] In principle, similar analyses can be performed on cued recall or free recall tasks by using Type II signal detection methods in which participants are given the option to withhold responses (Goldsmith & Koriat, 2007; Higham, 2002; Higham & Tam, 2005; Koriat & Goldsmith, 1994). However, there are a substantially smaller number of experiments in the existing literature for which these analyses are available or could be performed, making such designs less suitable for a meta-analysis.

vary across studies or conditions. This broad examination of potential influences on the aging of recognition memory, which is possible given the large amount of data included in the meta-analysis, allows us both to evaluate extant theories as well to test for potentially overlooked age differences that may point the way to future theoretical and empirical work.

Method

Selection of Studies

A computer-based search was conducted of the literature through the end of 2017 via PsycINFO using the keywords *recognition AND (aging OR ageing)*. In addition, manual searches were conducted via the reference lists of review and other articles and via prominent journals in the field (*Acta Psychologica*; *Aging, Neuropsychology, and Cognition*; *Brain and Cognition*; *Cerebral Cortex*; *Cognitive and Behavioral Neurology*; *Consciousness and Cognition*; *Developmental Psychology*; *Experimental Aging Research*; *Journal of Cognitive Neuroscience*; *Journal of Experimental Psychology: General*; *Journal of Experimental Psychology: Learning, Memory, and Cognition*; *The Journals of Gerontology: Series B*; *Journal of Memory and Language*; *Memory*; *Memory & Cognition*; *Neurobiology of Aging*; *Neuropsychologica*; *Neuropsychology*; *Psychological Science*; *Psychology and Aging*; and *Psychonomic Bulletin & Review*).

Experiments had to meet nine criteria for inclusion. If only some experiments within a multi-experiment paper met the inclusion criteria, we included those experiments and excluded the ones that did not meet the inclusion criteria. All of these criteria were determined prior to examining the results from the included (or excluded) experiments.

First, we included only papers in English.

Second, experiments needed to include a cross-sectional design with both a group of young adults (with a mean age no greater than 30) and a group of older adults (with a mean age no lower than 60). Occasionally, studies also included middle-aged groups; these samples were too infrequently

included for meta-analysis and were excluded.

Third, because the process of theoretical interest was healthy aging, we excluded clinical samples. When studies included both clinical and non-clinical samples, we did include the non-clinical sample.

Fourth, we included studies only in which young and older adults experienced the same experimental procedure. We excluded experiments in which the procedure varied across age groups (e.g., an experiment where older adults saw each item twice during the study phase, but younger adults saw each item only once) because in these cases it is ambiguous whether any group difference should be attributed to age or to procedural differences.

Fifth, because we were interested in natural age differences (if any) in recognition memory, the young and older adults had to be allowed to freely vary in their recognition memory performance. We excluded studies in which the inclusion of young and older adults was restricted so that the samples were deliberately matched in item discrimination or criterion (e.g., with the goal of examining age differences in neural activity or source memory when item memory was equated) because these studies do not provide information for or against what age differences might exist in an unconstrained sample.

Sixth, because we were interested both in memory fidelity and in remembers' strategic decisions concerning the remembering process, sufficient measures had to be presented to calculate both d' and $c,$ our indices of those two properties. (Typically, the presented measures were a hit rate and a false alarm rate, but d' and c could also be calculated from other pairs of measures, such as a measure of hits and a measure of hits minus false alarms.) When the relevant measures were presented only in bar graphs, we obtained numerical data by rasterizing the image and using Adobe Photoshop to measure the heights of the bars relative to the scale of the axes. We excluded studies that reported only hit rate or only hits minus false alarms; these data did not permit discrimination to be distinguished from criterion placement. We similarly excluded experiments in which the test phase consisted of only

targets or of only lures because these designs did not allow discrimination to be distinguished from criterion placement; we also excluded experiments in which subjects were explicitly instructed to affirm a fixed number of probes because these studies required subjects to set a particular criterion.

Seventh, we included in the meta-analysis only those studies that could be clearly coded for the independent variables of interest, which are detailed below. We excluded studies that did not report sufficient information to code all of these variables and studies in which the method fell outside the relevant categories for each variable, as discussed below for each individually (e.g., there were too few studies of age-differences olfactory recognition for meta-analysis).

Eighth, we included only old/new recognition tasks. Forced choice tasks, in which subjects must decide which of several items is the previously studied one, do not provide information about subjects' criterion for responding *old* or *new* because each trial typically contains exactly one old item.

Ninth, our theoretical interest was in simple memory for events, so we included only tasks in which subjects judged particular items as studied or unstudied; we excluded source memory tasks in which subjects were given an item known to be studied and judged which of several sources it came from. We also excluded tasks in which the item and source memory judgments were integrated and could not be separated, such as exclusion tasks, in which subjects are to affirm items from one studied list or source but reject new items and items from another list or source. Other meta-analyses have been conducted on source memory and related tasks, and age-related deficits on those tasks are widespread and well accepted (Old & Naveh-Benjamin, 2008; Spencer & Raz, 1995). We did include item memory data from experiments that included separate, consecutive item and source memory judgments (e.g., when a separate source memory judgment was made for each item judged as previously studied), and from inclusion task conditions of experiments that also had exclusion task conditions. We also included measures of old/new discrimination from experiments in which an item and source judgment were made simultaneously but separably (e.g., judging an item as being from

Source 1, from Source 2, or new; in this case, the first two categories represent an *old* response and the

lattermost a *new* response). Although our interest was in the behavioral *old*/*new* discrimination, we

included behavioral data from neuroimaging experiments as long as they met the other inclusion

criteria.

These criteria resulted in the inclusion of 501 experimental conditions from 232 experiments in

188 publications (approximate[2] total young adult N = 8,615; approximate total older adult N = 8,833).

See Figure 1 for the PRISMA diagram (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group,

2009) depicting the flow of publications through the stages of literature review. The complete dataset

and the R scripts for the statistical analyses and figures are available at

http://github.com/sfraundorf/FraundorfHourihanBenjamin_MetaAnalysis

---

[2] *N*s are approximate because some papers reported the total number of participants but did not report
the number of participants assigned to a between-subject condition used in the meta-analysis versus a
between-subject condition not used in the meta-analysis (e.g., the paper reported the total number of
participants, but did not report the number assigned to an inclusion task condition versus an exclusion
task condition). In these cases, we approximated the *N* per between-subject condition by dividing the
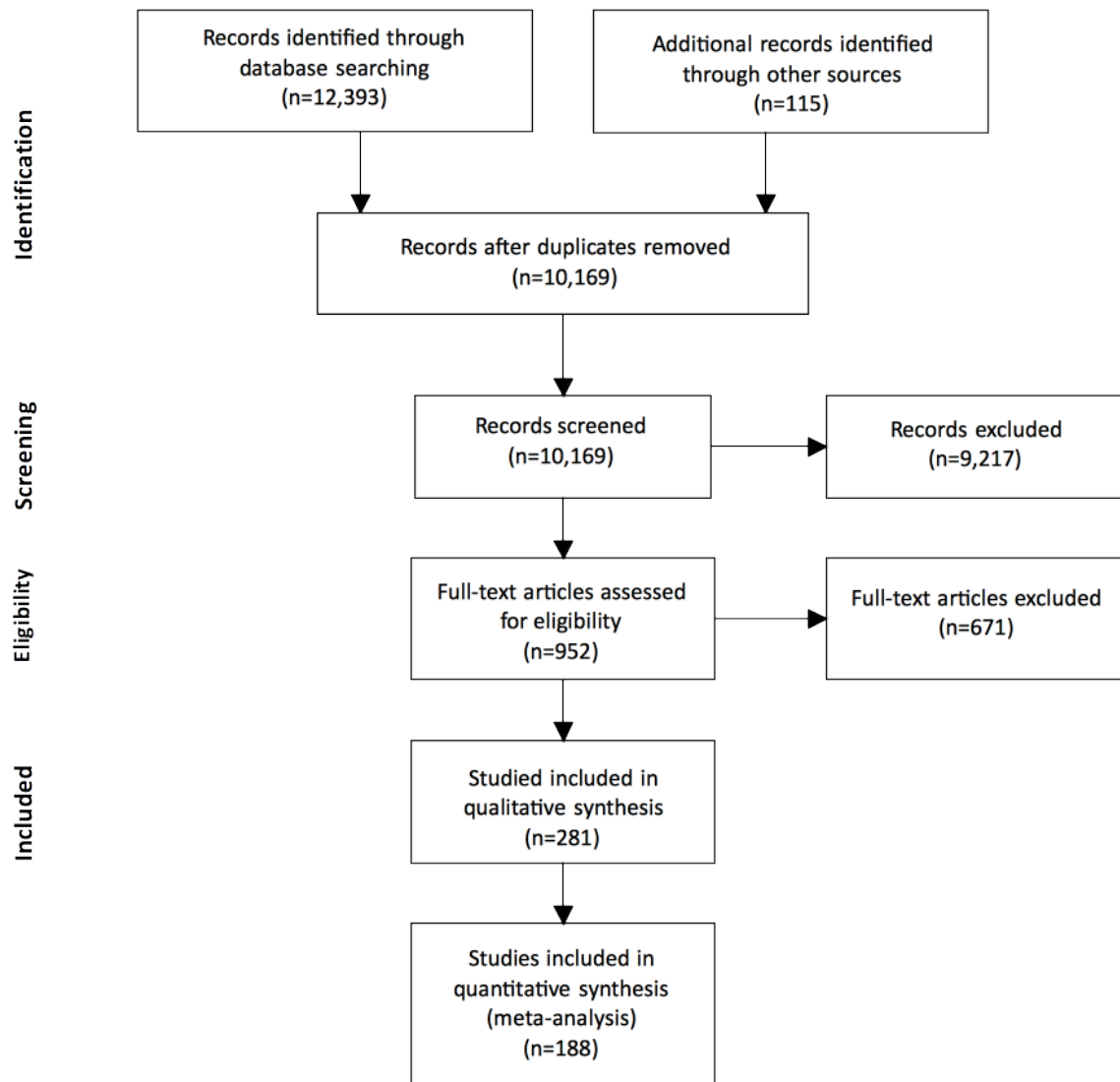total *N* among the number of between-subject conditions.

```
Identification

    ┌─────────────────────────┐      ┌─────────────────────────┐
    │  Records identified     │      │  Additional records     │
    │  through database       │      │  identified through     │
    │  searching              │      │  other sources          │
    │  (n=12,393)             │      │  (n=115)                │
    └─────────────────────────┘      └─────────────────────────┘
                │                                  │
                └──────────────┬───────────────────┘
                               ▼
                    ┌─────────────────────────┐
                    │ Records after duplicates │
                    │ removed                  │
                    │ (n=10,169)               │
                    └─────────────────────────┘
```

Screening

```
                    ┌─────────────────────┐       ┌──────────────────────┐
                    │ Records screened    │──────▶│ Records excluded     │
                    │ (n=10,169)          │       │ (n=9,217)            │
                    └─────────────────────┘       └──────────────────────┘
```

Eligibility

```
                    ┌─────────────────────┐       ┌──────────────────────┐
                    │ Full-text articles  │──────▶│ Full-text articles   │
                    │ assessed for        │       │ excluded             │
                    │ eligibility         │       │ (n=671)              │
                    │ (n=952)             │       └──────────────────────┘
                    └─────────────────────┘
```

Included

```
                    ┌─────────────────────┐
                    │ Studied included in │
                    │ qualitative         │
                    │ synthesis           │
                    │ (n=281)             │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │ Studies included in │
                    │ quantitative        │
                    │ synthesis           │
                    │ (meta-analysis)     │
                    │ (n=188)             │
                    └─────────────────────┘
```

Figure 1. PRISMA flow diagram of the literature review process.

Coding

Dependent measures. The dependent measures of interest were group-level age differences in the ability to discriminate old versus new items in recognition memory and group-level age differences in the bias (if any) to respond *old* versus *new*. Table 1 summarizes these measures. (We also conducted a supplementary analysis of hit rates and false alarms, available in Appendix A, which largely confirmed the results of the main analysis.)

We used *d'* and *c* as indices of discrimination[3] and response criterion, respectively (Green & Swets, 1966; Macmillan & Creelman, 2005). Since most papers included in the meta-analysis did not report *d'* and *c* scores, we calculated them using the reported behavioral measures (e.g., hit rates and false alarm rates in each condition). *d'* is undefined given a perfect hit rate of 1 or a perfect false alarm rate of 0; in these cases, we replaced the proportion of 0 or 1 with a proportion equal to half of one trial (Macmillan & Creelman, 2005, p. 8).

For most papers, *d'* and *c* scores could only be calculated at the group level—that is, using the group mean hit rate and group mean false alarm rate rather than using individual hit and false alarm rates (and then averaging the resulting individual *d'* and *c* scores). However, such grouped data can often perform well[4] (A. Cohen, Sanborn, & Shiffrin, 2008; but see Estes & Maddox, 2005). Some papers did report the mean of *d'* and/or *c* scores that had been calculated at the individual level; we used such data when they were available[5].

In some experiments, multiple categories were available for old and/or new items, such as when subjects responded as *remember*, *know*, or *new*, or when subjects judged each probe as *sure old*, *probably old*, *probably new*, or *sure new*. The selection of such categories varied from experiment to experiment, so to facilitate comparison we summed the responses in each case to create a single *old* category and a single *new* category and then calculated *d'* and *c*.

---

[3] Another measure of discrimination is a nonparametric measure of area under the curve, A_z (Stanislaw & Todorov, 1999). However, given only a single hit rate per condition (i.e., in the absence of confidence ratings), between-person differences in *d'* and Az will correlate highly. Indeed, in our dataset, A_z correlated with *d'* at .93 and led to virtually the same conclusions.

[4] Moreover, even participant-level analyses essentially constitute grouped data because they aggregate over multiple items judged by the participant. (For further discussion, see Pratte & Rouder, 2012.)

[5] In cases where the average individual *d'* score was available, but not the average individual *c* score, we used the group-level *d'* score so that the *d'* score was based on the same data, and could be compared to, the group-level *c* score.

Table 1

Mean and Standard Deviation of Dependent Measures Across Experimental Conditions Included in Meta-Analysis.

Fixed effect	M	SD
d' (young adults)	1.85	0.80
d' (older adults)	1.39	0.71
Age group difference in d'	0.46	0.42
c (young adults)	0.16	0.29
c (older adults)	0.13	0.34
Age group difference in c	0.04	0.21

Note. SD = standard deviation. The mean age group differences may not exactly equal the difference in age means because of rounding.

Independent measures. One of the goals of the study was to test the influence of variables that have been claimed to moderate age differences in memory. We selected *a priori* a set of variables that were suggested by the literature as important in characterizing the subjects, the to-be-remembered materials, and the study, retention, and test procedures and that could be obtained from most published reports, and we coded each study on these dimensions. Table 2 presents a summary of the independent measures.

Table 2

Mean and (Where Applicable) Standard Deviation and Frequency Counts of Independent Measures Across Experimental Conditions Included in Meta-Analysis.

Fixed effect	M	SD	Count
Subject characteristics			
Mean age of older adult group (years)	71.2	3.3	—
Stimulus characteristics			
Stimuli are words	52.1%	—	261
Stimuli are pictures	27.1%	—	136
Stimuli are faces	14.0%	—	70
Stimuli are texts	8.6%	—	43
Positive valence	6.8%	—	34
Neutral valence	85.4%	—	428
Negative valence	7.8%	—	39
Study phase characteristics			
Visual-only presentation	85.8%	—	430
Auditory-only presentation	4.8%	—	24
Bimodal (visual-auditory) presentation	9.4%	—	47
Production task	9.2%	—	46
Intentional encoding	64.9%	—	325
Self-paced study	8.2%	—	41
Study time per presentation when not self-paced (s)	3.7	2.6	—
Deep orienting task	27.3%	—	137
Shallow orienting task	8.8%	—	44
Generation task	1.6%	—	8
Divided attention at study	0.8%	—	4
Number of to-be-remembered items	83.6	72.6	—
Multiple study opportunities per item	7.6%	—	38
Test phase characteristics			
Intervening cued recall	1.2%	—	6
Intervening free recall	7.6%	—	38
Continuous recognition task	4.6%	—	23
Retention interval (log min.)	1.1[a]	18.4	—
Test list length (items)	112.8	107.6	—
Conjunction lures	9.4%	—	47
Component lures	14.2%	—	71
Semantically-related lures	6.8%	—	34
Featurally-related lures	4.0%	—	20
Unrelated lures	70.9%	—	356
Proportion of lures	.50	.10	—

Note. SD = standard deviation. Categorical variables do not sum to 100% because multiple codes can apply to the same condition (e.g., lures can be both semantically related and featurally related). [a]A retention interval of 1.1 log-minutes equals approximately 3.0 minutes.

Older adult age mean. We obtained the mean age (in years) of the older subject group from each paper. When only an age range was reported and no mean age, we used the midpoint of the age

range. We did not include the mean age of the young adults (or the difference in mean age between the young and older adults) because the homogeneity of most young adult samples (typically college students between 18 and 22) meant that there was little variance in this variable.

Stimulus type. We coded the presence or absence of each of four types of study materials: *words* (single words and word pairs in the subjects' native language), *faces, pictures* (whether drawn or photographs), and *text* (longer, semantically meaningful verbal materials, such as sentences or paragraphs).   We also report additional analyses of the distinction between faces and other stimuli because it has been argued that faces are processed differently at both a neural level (Kanwisher, McDermott, & Chun, 1997) and cognitive level (e.g. potentially by requiring more holistic processing; for further discussion, see Tanaka & Simonyi, 2016).

If a particular memorandum consisted of two stimulus types presented together, such as a picture presented alongside a verbal label, both of the applicable codes were applied. Other types of stimulus materials, such as odors, motor actions, proper names, music, videos, and foreign vocabulary words, were not present in a sufficient number of experiments to be analyzed and were excluded from the meta-analysis. We also excluded cases in which data were reported collapsing over multiple stimulus types (e.g., each item was either a picture alone or a word alone, but data were not presented separately for pictures versus words).

Affective valence. We coded whether the affective valence of the stimuli was *positive*, *negative*, or *neutral*. Negatively-valenced stimuli included words, texts, and pictures that were rated as negatively valenced on existing norms or that were explicitly reported by the authors to be negative, as well as faces expressing anger, sadness, or fear. Positively-valenced stimuli included words, texts, and pictures normed or explicitly reported to be positive, as well as faces expressing happiness. All other stimuli (e.g., lists of words or concrete objects), as well as those explicitly reported by the authors to be affectively neutral, were assumed to be neutral.

We excluded from the meta-analysis experiments containing items of mixed valence (e.g., a mix of positive and negative stimuli) for which data were not reported separately for each condition. A small number of experiments manipulated the affective valence separately during the study and test phases (e.g., a face was presented with a happy expression during the study phase but an angry expression during the test phase). Because such experiments were too infrequent to separately examine effects of valence at study from effects of valence at test, and it was not clear a priori which would be more important, we excluded conditions in which the valence of an individual item changed from study to test.

Because in most cases the valence of the materials was clearly reported in the original publications, we expected it could be coded with high reliability across raters. To confirm this, a second rater judge the stimulus valence of a randomly chosen 10% of the experiments included in the analysis. Agreement was approximately 95%, suggesting that this was indeed a reliable judgment.

Study modality. We coded study modality as *visual*, as *auditory*, or as *bimodal* if materials were presented in both modalities. Other modalities examined in only a few studies (e.g., haptic or olfactory perception) were excluded.

Production. Tasks were coded as involving *production* if subjects were required during encoding to produce the to-be-remembered stimuli by saying them out loud or writing them down or were coded as involving *no production* if subjects simply viewed or heard the items silently.

Intentionality of encoding. Procedures were coded as *intentional* if subjects were informed that their memory for the materials would be tested and as *incidental* if subjects were not so informed.

Study time. Presentation was coded either as *self-paced* if subjects could control the time for which each memorandum was presented or as *experimenter-paced* if the presentation duration was fixed. For the experimenter-paced, fixed-duration conditions, we also obtained the presentation

duration in seconds (per presentation, in the case of multiple presentations of a single item). In the infrequent situation that presentation duration varied across items (e.g., spoken sentences that varied in their length), the mean presentation duration was used. Presentation duration was coded only for experimenter-paced conditions because few studies reported the average length for which subjects studied the memoranda in self-paced conditions, and such averages would likely be misleading anyway because self-pacing behavior is highly variable across individual subjects (Tullis & Benjamin, 2011).

The self-pacing code encompassed both instances of self-paced deliberate study and instances where subjects controlled the presentation rate through the speed at which they completed an incidental task (e.g., each presentation lasted exactly as long as it took for subjects to rate the pleasantness of that particular item). However, it is plausible that age differences in the effects of self-pacing on memory could emerge only for deliberate self-paced study. Thus, we also included an interaction between self-pacing and intentionality of encoding to capture any effects that emerged specifically in instances of self-paced deliberate study.

Task at study. Some experiments involved a concurrent task during the study phase. We coded each of these tasks for the presence or absence of each of four characteristics described below. Because this coding procedure was more complex than the procedures for most of the other variables (e.g., coding the presentation modality), we assessed the reliability of the study task coding procedure by having the first two authors independently categorize each task. On each of the four task characteristics, agreement between the two raters (as measured by Cohen's kappa) was substantial according to the criteria of Landis and Koch (1977); disagreements were resolved through discussion.

Tasks that involved *deep orienting* ($\kappa = .70$) asked subjects to perform a judgment related to meaning or to individuating features of a face, such as judging whether a noun represented an animate object or deciding on a face's most distinctive feature. Tasks that involved *shallow orienting* ($\kappa = .62$) asked subjects to perform a judgment related to visual, orthographic, or phonological properties, such

as counting the syllables in a word or judging whether a face is attractive. Tasks that involved *generation* (κ = .72) required the subject to perform some processing on a cue to obtain the to-be-remembered stimuli, such as completing a word stem or thinking of the antonym of a presented word. Finally, tasks that involved *divided attention* (κ = .81) required the subject to perform a concurrent task on stimuli unrelated to the to-be-remembered items, such as performing a digit span task while studying a list of to-be-remembered words.

Note that there were some experimental conditions for which none of these characteristics applied; indeed, most conditions simply presented stimuli without any concurrent task whatsoever. Conversely, some conditions were coded as having more than one of the above characteristics (e.g., subjects performed shallow orienting task on the to-be-remembered stimuli while also having their attention divided by a concurrent digit monitoring task). We excluded from the meta-analysis studies in which multiple different tasks were presented and data were not presented separately for each task type (e.g., some items were presented with a deep orienting task whereas others presented with a shallow orienting task, but data were presented only collapsing across the task types).

Number of memoranda. We coded the number of items to which subjects were exposed in the study list. In cases of multiple cycles of study phase and test phases, we counted the number of items per study list. Some experiments included "buffer" items at the beginning and end of the study list that were not actually analyzed; we included these in the count of memoranda because, from the subject's perspective, these were still to-be-remembered items and nothing explicitly indicated that the researchers would not analyze them.

Number of study opportunities. We coded whether each memorandum was studied *one* or *more than one* time. Most experiments involved only single presentations, and there was little variance in the number of presentations among those studies that included multiple presentations, so we did not attempt to distinguish among different numbers of multiple presentations.

Intervening recall test. We coded whether or not each of two types of recall test—*cued recall* and *free recall*—was presented before the recognition test. (In some cases, both types of recall test were presented before.) We excluded from the meta-analysis experiments in which a recall task sometimes preceded recognition test and sometimes did not (i.e., if the order of the recall and recognition task was counterbalanced) and data were not presented separately with and without the intervening recall.

Task type. We coded the recognition task as *delayed recognition* if no test was presented until the study phase was concluded or as *continuous recognition* if each test probe was also a to-be-remembered stimulus.

Retention interval. We coded the time in minutes that elapsed between when an item was studied and when the memory test occurred. However, this variable was necessarily defined slightly differently for continuous recognition tests than for delayed recognition tests. For delayed recognition tests, retention interval was defined as the time between the end of the study phase and the beginning of the first test. However, for *continuous recognition* tests, in which each test probe was also a to-be-remembered stimulus, there was no distinction between study and test phases; thus, retention interval was defined as the time between the study presentation and the test presentation. To accommodate the different definitions entering into the retention interval variable, we allowed retention interval to interact with the continuous recognition variable in determining age differences.

Because memory typically declines over a retention interval according to a power law (Wixted, 2004), in which memory declines more quickly immediately following learning than later, retention interval was log-transformed before being entered into the model. Tests that immediately followed study without any delay or intervening task were assigned the minimum retention interval of 0.08

minutes[6].

     Test list length. We coded the number of test probes (including both targets and lures) presented in the test phase (or per test list in the case of multiple study-test cycles). As with the number of memoranda, we included unanalyzed buffer items in this count since they were not discernably different to the subject.

     Lure type. The new, unstudied lures presented during the test phase were coded for whether and how they related to the studied items. As with the study task coding, to assess the reliability of this more complex coding procedure, we had the first two authors independently categorize each task. On each of the four task characteristics, agreement between the two raters was almost perfect according to the criteria of Landis and Koch (1977); disagreements were resolved through discussion.

     Conjunction lures ($\kappa = .96$) consisted entirely of parts of previously presented items recombined to create a new item, such as the two words that comprised a word pair re-paired to create a new pair or a face and hat presented together during the study phase re-paired to create a new face-hat pairing. *Component* lures ($\kappa = .92$) consisted of part of a previously presented item combined with a new, unstudied part, such as one word from a previously studied word pair now paired with a new, unseen word or a studied face presented with a previously unseen hat to create a new face-hat pairing. *Semantically related* lures ($\kappa = .85$) were wholly or partially new items that related to particular studied items at the meaning or conceptual level. *Featurally related* lures ($\kappa = .83$) were related to particular studied items at the level of surface similarity (e.g., words with similar spelling or faces with similar noses). Finally, lures coded as *unrelated* did not have any relation to particular targets aside from being drawn from the same broad population of possible items (e.g., English nouns).

     As with the orienting task classification, more than one of these characteristics could be applied

---

[6] We did not assign these instances a retention interval of 0 because the log transformation is undefined for values of 0; instead, we used the smallest value otherwise observed in the dataset.

to a single condition if the lures simultaneously related to particular targets in multiple ways (e.g., a lure word that both belonged to the semantic category as a target word *and* rhymed with that target word would be coded as having both a semantic and a featural relation). However, we excluded from the meta-analysis studies in which multiple different lure types were presented and data were not reported separately for each type.

Proportion of lures. We coded the proportion of test items that were new/unstudied and should thus be rejected. Note that as the proportion of lures rises, so too does the optimal criterion for terming an item *old/studied* (i.e., subjects should be more conservative when there are more lures). Since it is implausible for a task to contain 0 percent lures, we centered the proportion-of-lures variable around 0.5 so that a proportion-of-values of 0 corresponded to an equal ratio of targets and lures in order to facilitate interpretation.

Other variables. Some experiments also reported data divided by variables other than the ones described here, such as whether a visually presented image depicted a physically possible or impossible figure. These variables were not included in the meta-analysis because they were manipulated in too few studies and/or did not widely figure into prominent theoretical accounts of age differences in memory. When experiments included conditions divided by these other variables (e.g., a physically possible image condition and a physically impossible image condition), we pooled those conditions for the meta-analysis.

Analytic Procedure

As noted above, we were interested in (a) whether there were overall mean differences between young and older adults in recognition memory discrimination (as described by d') and response criterion (as described by c), (b) whether age differences increased or decreased with increasing d' and c (e.g., were older adults especially disadvantaged relative to younger adults on more difficult tasks?), and (c) whether these differences were moderated by the stimulus, study phase, and test phase

characteristics described above.

We conducted our meta-analysis within the framework of linear mixed-effects regression (Baayen, Davidson, & Bates, 2008; for applications to meta-analysis, Stram, 1996). Because the dataset included only studies in which young and older adults were tested on the same procedure, each experimental condition constituted a pair of d' data points (one d' for young adults and one for older adults) and a pair of c data points, allowing one difference score to be calculated for d' and one for c. We thus conducted two regressions. In one regression, we treated age differences in recognition memory discrimination as the outcome measure and their d' measures in each condition in each study as the to-be-predicted individual data points. Specifically, for every measure of older adults' d' in a particular condition, the regression model included three predictors. First, corresponding to research question (a) above, we included an intercept term that captured any mean age difference in recognition memory discrimination. Second, corresponding to research question (b), we included the younger adults' d' in the same condition; this measure was mean-centered so that the intercept term corresponded to the average age difference in d'. A slope of 1 would indicate that age differences were invariant across the range of task difficulty, a slope greater than 1 would indicate that age differences became smaller (i.e., older adults were relatively less impaired) on tasks that were easier for younger adults, and a slope less than 1 would indicate that age differences were *larger* (i.e., older adults more impaired) on easier tasks. Third, corresponding to research question (c), we included predictor variables for the stimulus and task characteristics described above[7]. These variables tested whether characteristics of the stimuli or of the experimental procedures led to age differences above and beyond those predicted from the mean age difference and from younger adults' performance. For example, if

---

[7] In principle, it would be possible to also include all of the interactions among the variables (e.g., a Generation x Valence interaction to test whether age differences in the generation effect are larger for negative stimuli than neutral stimuli). However, with 29 predictor variables, this would result in a combinatorial explosion of interactions, which cannot be feasibly modeled and would lead to excessive Type I error. Rather, we examined only those interactions for which we had a priori hypotheses.

older adults have a special deficit in associative memory (Old & Naveh-Benjamin, 2008), tasks that involve conjunction lures should engender particularly low performance among older adults, beyond that predicted from overall age differences and task difficulty alone. Using these predictor variables, we related age differences in d' to younger adults' d' and to the task characteristics; in a second regression, we conducted an analogous regression for the measure of response criterion, c.

In addition to these *fixed effects* of theoretical interest, both models included several random effects to capture between-experiment variability. Even close replications of the same experimental procedure can differ substantially in the effect size yielded, perhaps because of differences in subject populations or unreported methodological differences (*method factors*; McShane & Böckenholt, 2014). Failing to account for these differences inflates the Type I error rate (Hunter & Schmidt, 2000). We controlled for these sources of variability by including them as *random effects*, effects for which the individual categories are sampled out of a larger population (e.g., the subjects in individual experiments are sampled out of a population). Most typically, random-effects meta-analyses allow for variability of effect sizes across individual experiments (e.g., to account for methodological differences). In the present dataset, additional sources of variability (Hedges, Tipton, & Johnson, 2010; Konstantopoulos, 2011) include (a) superordinate method factors common across all of the experiments within a multi-experiment publication (of which many were included in the meta-analysis), (b) superordinate method factors common across all of the experiments from the same laboratory (e.g., characteristics of the local subject population), and (c) subject-level variability, as reflected in the fact that in some experiments the same group of subjects contributed data to multiple conditions (i.e., within-subject designs). Thus, we included as random effects (a) publication, (b) laboratory, (c) subject groups, and (d) experiments. For each of these random effects, we included a *random intercept* to account for additional variability (across subjects, experiments, and publications) in d' and in c.

Sample size varied across the experiments included in the meta-analysis. All other things being

equal, a study that included more participants provides more precise information and should be given greater weight in the meta-analysis; thus, in the random-effects regression, we weighted each study proportionate to its sample size (Hunter & Schmidt, 2004). (In principle, an alternative proposed by Hedges and Vevea [1998] is to weight experiments according to the inverse of their variance, such that studies that provide more precise estimates are given greater weight; however, the majority of papers reporting recognition memory experiments do not include estimates of the across-subject standard deviations or variance in d' and c.)

Finally, we evaluated the statistical evidence for or against the presence of publication bias, namely, a bias for statistically significant comparisons to be published over non-significant comparisons. Although there exist numerous methods for assessing or correcting for publication bias (for recent reviews, see McShane, Böckenholt, & Hansen, 2016; Carter, Schönbrodt, Gervais, & Hilgard, 2017), many require the original papers to report p-values or estimates of the between-participant variance in effect size. Unfortunately, that information was not available for the majority of the papers included in the present meta-analysis; for instance, few original papers actually report comparisons of young and older adults in their criterion placement. But, assessment of publication bias can also be performed by examining the relation between sample size and the estimated effect size (J.L. Peters, Sutton, Jones, Abrams, & Rushton, 2006). If there is no publication bias, these quantities should show no linear relationship. But, in the presence of a bias to publish statistically significant results, small-sample studies that happen to find a large effect size (due to their sampling error) are more likely to be published whereas small-sample studies that find a spuriously small and non-significant effect are not. This problem might be particular acute in studies of memory aging because older participants may be more difficult to recruit. Thus, we assessed publication bias by testing whether (the inverse) sample size predicted the size of the age difference in d' and/or in c (in the presence of all other between-study variables, as is appropriate for datasets with between-experiment

heterogeneity, such as this one; J.L. Peters, Sutton, Jones, Abrams, Rushton, & Moreno, 2010).

All models were fit in the R Project for Statistical Computing using the *lmer*() function of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014). (The exact model specifications are available in Appendix B.) We assessed the statistical significance of each variable of interest using likelihood-ratio tests, which avoid anticonservativity (Barr, Levy, Scheepers, & Tily, 2013). Following Baayen (2008, p. 270), 95% confidence intervals for the age differences were constructed from the parameter estimate plus or minus two times the standard error[8].

Results

Discrimination

We first examine age differences in discrimination; that is, the ability to discern which items were previously studied and which were not. Table 3 displays the results from the model of age differences in *d'* discrimination scores.

---

[8] This procedure reflects the fact that, for mixed effect models with hundreds of degrees of freedom, the t-distribution has essentially converged to the normal distribution, but uses the slightly more conservative threshold of 2.0 times the standard error (rather than 1.96) to account for the fact that the *t*-distribution never *exactly* converges to the normal distribution (Baayen, 2008).

Table 3
Fixed Effect Estimates for Multi-Level Model of Differences Between Age Groups in d' *Scores.*

Fixed effect	$\hat{\beta}$	SE	95% CI	χ^2	p
Intercept (baseline age difference)	0.463	0.026	[0.409, 0.513]		
Younger adult *d'* score (slope)	0.257	0.023	[0.212, 0.301]	117.79	< .001
Subject characteristics					
Mean age of older adult group	0.006	0.007	[-0.007, 0.019]	0.88	.35
Stimulus characteristics					
Stimuli are faces (vs. words)	0.141	0.062	[0.019, 0.263]	5.41	.02
Stimuli are pictures (vs. words)	-0.071	0.044	[-0.156, 0.015]	2.48	.12
Stimuli are texts (vs. words)	-0.208	0.066	[-0.338, -0.078]	9.25	< .01
Positive valence	0.064	0.052	[-0.037, 0.165]	1.63	.20
Negative valence	0.080	0.047	[-0.012, 0.171]	3.05	.08
Study phase characteristics					
Auditory presentation (vs. visual)	0.094	0.074	[-0.126, 0.469]	1.59	.21
Bimodal presentation	0.172	0.152	[-0.126, 0.469]	1.40	.24
Production	-0.164	0.153	[-0.464, 0.135]	1.28	.26
Intentional encoding	0.046	0.054	[-0.061, 0.152]	0.77	.38
Self-paced study	0.187	0.086	[0.018, 0.356]	5.11	.02
Self-paced study x intentional encoding	-0.007	0.131	[-0.264, 0.251]	< 0.01	.98
Study time per presentation (s)	0.012	0.009	[-0.005, 0.029]	2.01	.16
Deep orienting task	0.111	0.051	[0.010, 0.211]	4.99	.03
Shallow orienting task	0.066	0.065	[-0.061, 0.194]	1.24	.27
Generation task	-0.139	0.094	[-0.323, 0.045]	2.35	.13
Divided attention	-0.127	0.187	[-0.493, 0.240]	0.47	.49
Number of memoranda	-0.001	0.001	[-0.001, 0.000]	1.90	.17
Multiple study opportunities	-0.070	0.058	[-0.184, 0.043]	1.54	.21
Test phase characteristics					
Intervening cued recall	0.196	0.143	[-0.083, 0.475]	1.99	.16
Intervening free recall	-0.196	0.085	[-0.363, -0.030]	5.63	.02
Continuous recognition	0.016	0.127	[-0.233, 0.265]	0.01	.91
Retention interval	0.036	0.007	[0.023, 0.049]	28.47	< .001
Continuous recognition x retention interval	0.033	0.031	[-0.029, 0.094]	1.13	.29
Number of test probes	< 0.001	< 0.001	[-0.001, 0.001]	0.04	.85
Conjunction lures	0.333	0.046	[0.244, 0.423]	53.40	< .001
Component lures	0.144	0.068	[0.010, 0.278]	4.65	.03
Semantically related lures	0.241	0.037	[0.169, 0.313]	42.70	< .001
Featurally related lures	0.037	0.047	[-0.055, 0.129]	0.65	.42
Proportion of lures	-0.117	0.172	[-0.454, 0.220]	0.46	.50

Note. SE = standard error. Positive parameter estimates indicate larger age differences (i.e., older adults especially disadvantaged relative to young adults); negative parameter estimates indicate age differences smaller than the mean (i.e., older adults were relatively less disadvantaged).

Overall performance. Older adults had lower *d'* scores than younger adults. The model

estimated the overall age difference as 0.46 *d'* units (95%: CI [0.41, 0.51]); that is, when young adult

performance was at its mean ($d' = 1.85$), older adults' discrimination was 0.46 d' units worse[9], close to

a medium effect size according to the standards of J. Cohen (1988). Given younger adults' average

level of performance ($d' = 1.85$), this d' difference translates into a decrease of approximately 7% in

the hit rate with a simultaneous 7% increase in the false alarm rate under an unbiased criterion. This

overall age difference in discrimination is captured by the significant intercept term[10], $t = 18.32$, $p <$

.001.

Further, this difference was more pronounced in easier tasks. The slope relating younger adults'

d' score to the age difference in d' scores was approximately 0.26 (95% CI: [0.21, 0.30]); thus, for

every increase of 1 d' unit in younger adults' memory performance, the deficit in older adults'

performance relative to young adults' grew by an additional 0.26 d' units. Put another way, for every

increase of 1 d' unit in younger adults' memory performance, older adults gained only 0.74 of a d' unit;

or, if younger adults increased their performance by 1 standard deviation (from $d' = 1.46$ to $d' = 2.25$),

they would improve their performance with a 10% increase in hit rate and 10% decrease in false alarm

rate whereas older adults would show only an 8% change in each of these measures.

These relationships are depicted in Figure 2, which plots the group-level d' scores for older

adults as a function of those for younger adults. If there were no age differences in recognition

memory, all points would lie along the diagonal line. Instead, the vast majority of points are below the

---

[9] One potential concern is that this analysis might be unduly influenced by outlying results from one or two experiments that adopted unusual procedures. To address this concern, we also considered a version of the model in which we removed observations with standardized residuals more than 3 standard deviations from the mean; these observations are those that are outliers after accounting for all of the fixed and random effects (Baayen, 2008, p. 207).   If we excluded the 1% of data that were outlying under this criterion, the only change was that (a) the effect of pictures in reducing age differences in memory was now significant at the $\alpha=.05$ level and (b) there were now only marginal effect of self-paced study ($p = .06$) and of deep encoding ($p = .07$). This pattern suggests that a beneficial effect of pictures for older adults characterized the majority of the dataset but that there were a small number of exceptions.

[10] We assess the significance of the intercept using a t-test because the likelihood-ratio test is not available for tests of the intercept.

line; for almost every experiment and task condition, the group-level *d'* score for older adults was lower than the corresponding young adult score. (In the Discussion, we revisit the small number of points in which older adults outperformed young adults.) Moreover, this age deficit grew as young adults' discrimination performance increased (i.e., moving towards the right-hand side of the figure).
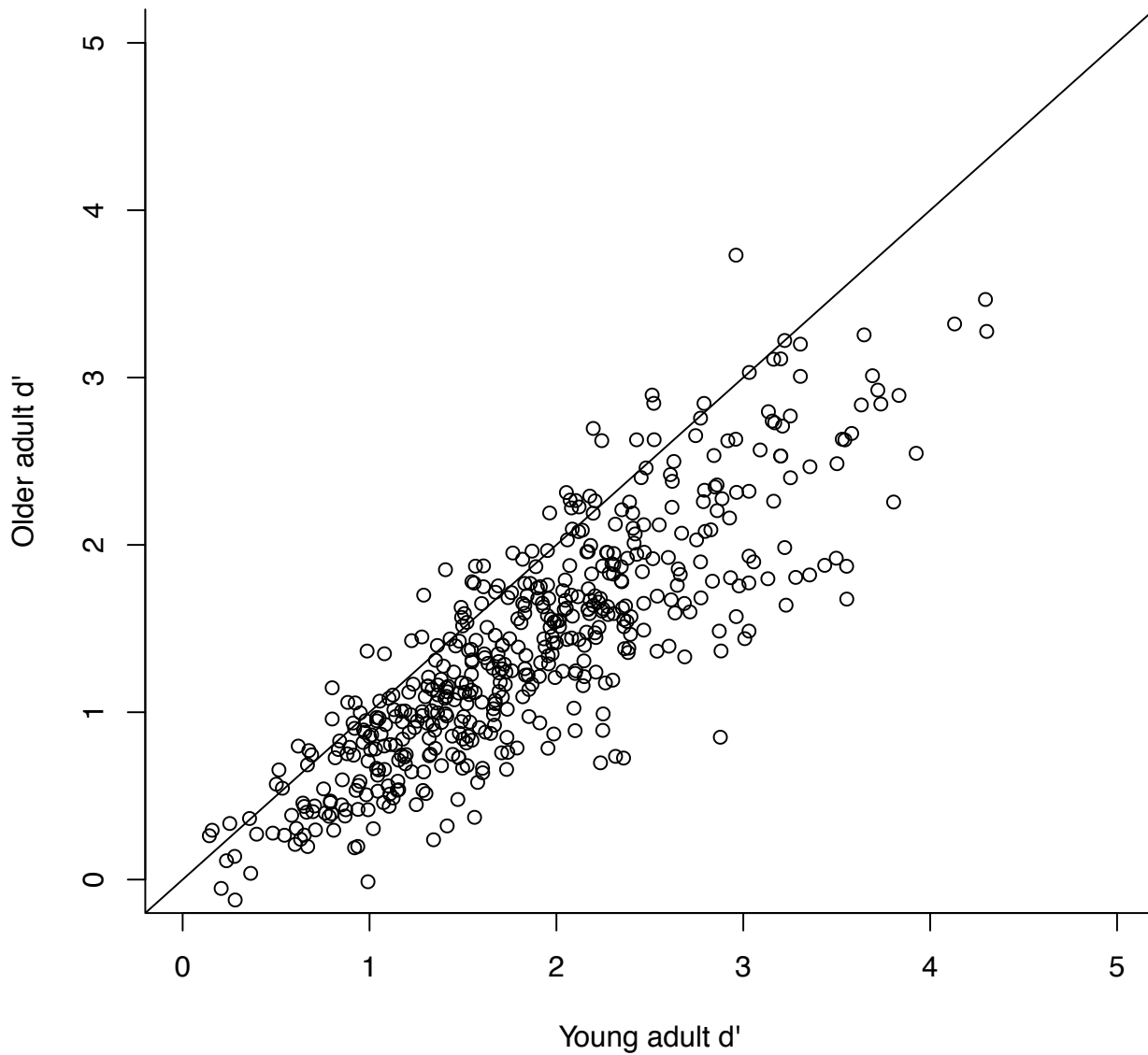
Figure 2. Group-level *d'* scores for older adults as a function of group-level *d'* scores for younger adults. Each point represents one experimental condition. The diagonal line (identity) represents the relation that would be obtained if there were no age differences in discrimination performance.

Task characteristics. However, age differences in *d'* scores were not wholly predicted by the

above linear relation to young adults' scores. Rather, in some experimental conditions, the age

difference was larger or smaller than what would be otherwise predicted by young adults' *d'* scores;

that is, there was heterogeneity across conditions. As a statistical test for this heterogeneity, we compared a simple model that modeled age differences with only a fixed intercept (i.e., a fixed-effects model) to a random-effects model that allowed the age difference to vary across conditions. The random-effects model fit significantly better, $\chi^2_{(1)} = 229.5$, $p < .001$, providing significant evidence for heterogeneity in the age difference across experimental conditions[11].

In light of this heterogeneity, it can be asked whether certain types of memory tasks consistently produced larger age differences than others. To answer this question, we return to our main model (as reported in Table 3), which tested whether age differences were modulated by certain features of the experimental conditions. This model indicated that older adults were especially disadvantaged by many conditions in which the new, unstudied lures bore special resemblance to the studied, target items. For example, when the lures were a novel conjunction of studied parts (e.g., two words re-paired to form a new pair), older adults were on average an *additional* 0.33 *d'* units worse than would be predicted from the linear relation alone. (That is, conjunction lures magnified the difference between young and older adults.) Age differences in discrimination performance were also amplified, though to a lesser degree, by lures that included one studied component paired with one novel component (0.14 additional *d'* units larger) and by lures that bore a semantic relationship to particular studied targets (0.24 additional *d'* units larger). However, lures that bore a surface-level featural resemblance to particular targets did not significantly magnify the age difference relative to unrelated lures.

Older adults' recognition memory was also especially disadvantaged for faces relative to words. Age differences in discrimination were also greater when subjects controlled the rate at which items were presented during study (regardless of whether this self-pacing occurred under intentional or

---

[11] Similar tests also revealed significant heterogeneity across publications, laboratories, and subject groups. The final model estimated the standard deviation across experiments to be 0.16 *d'* units, 0.16 across publications, 0.09 across laboratories, and 0.11 across subject groups within an experiment.

incidental encoding instructions) and when subjects were given an encoding task that emphasized "deep" processing (e.g., judging whether a noun represented an animated object or whether a particular picture was pleasant or unpleasant). Finally, despite an overall tendency for age differences in discrimination to be relatively smaller on harder tasks (as noted above), longer retention intervals between study and test were associated with larger age differences.

By contrast, older adults' recognition memory was relatively spared for texts (sentences or paragraphs). For texts, age differences in recognition memory were 0.21 *d'* units *smaller* than what would be predicted by the overall linear relation described above; that is, older adults were not as disadvantaged in remembering texts. (Interestingly, this change appears to be driven largely by hits to previously presented texts; see Appendix A.) Age differences in recognition memory were also smaller (by 0.19 *d'* units) when a free-recall test intervened between initial study and the recognition memory test. Note, however, that both of these parameter estimates are of smaller magnitude than the overall age difference in discrimination performance (0.41 *d'* units); that is, texts and intervening free recall tests *reduced* age-related deficits but did not *eliminate* them.

There was no significant evidence that older adults were differentially affected by presentation modality, self-generation at study, shallow encoding tasks, divided attention, intervening cued recall, list length at study or test, intentionality of encoding, or number of repetitions.

To assess whether these results may have been influenced by publication bias, we assessed whether inverse sample size predicted the reported size of the age difference in *d'* (J.L. Peters et al., 2010). Adding this variable to the model revealed that there was indeed evidence for publication bias insofar as larger effect sizes were observed in studies with smaller samples, $\chi^2_{(1)} = 6.81$, $p < .01$. There should be no such relation in an unbiased literature, and its presence suggests that other, small-sample studies that found smaller, non-significant effects may not have entered into the published literature. Nevertheless, even controlling for this publication bias, the overall age difference in *d'* remained

sizable, at approximately 0.34, as did the significant effects of all of the between-condition variables discussed above.

Lastly, to determine whether any of these effects differ for faces, since these stimuli may enjoy special status, we examined the interactions of the Face Stimuli variable with each other variable[12]. Only one such interaction emerged: Valence interacted with face stimuli such that age deficits were *reduced* given negatively-valenced faces than neutral faces, $t = -2.28$, $p = .02$. By contrast, for other stimuli, age deficits were *exaggerated* for negatively-valenced stimuli, $t = 2.16$, $p = .03$.

Criterion placement

We next turn to the criterion that the groups set for terming an item *studied/old*. Recall that a higher criterion indicates a conservative bias to judge items as unstudied/new whereas a lower criterion indicates a liberal bias to judge items as studied/old. Table 4 displays the results from the model of age differences in the placement of this criterion; positive estimates indicate variables that favored young adults setting a higher, more conservative criterion for terming an item *studied/old* relative to older adults (i.e., the young adults' group-level c is greater than the older adults' group-level c) whereas negative numbers conversely indicate variables that influenced older adults in the direction of setting a more conservative criterion relative to young adults.

---

[12] We could not perform this analysis for four variables—generation, divided attention, intervening cued recall tasks, and intervening free recall tasks—because those features never co-occurred with face stimuli.

Table 4
Fixed Effect Estimates for Multi-Level Model of Differences Between Age Groups in c *Scores.*

Fixed effect	$\hat{\beta}$	SE	95% CI	χ^2	p
Intercept (baseline age difference)	0.045	0.014	[0.018, 0.072]		
Younger adult c score (slope)	0.128	0.033	[0.064, 0.192]	15.53	< .001
Subject characteristics					
Mean age of older adult group	0.005	0.004	[-0.003, 0.012]	1.68	.19
Stimulus characteristics					
Stimuli are faces (vs. words)	0.213	0.036	[0.142, 0.284]	34.46	< .001
Stimuli are pictures (vs. words)	0.096	0.025	[0.047, 0.145]	13.89	< .001
Stimuli are texts (vs. words)	0.052	0.038	[-0.023, 0.126]	1.86	.17
Positive valence	0.026	0.032	[-0.036, 0.089]	0.63	.43
Negative valence	-0.008	0.029	[-0.064, 0.049]	0.10	.75
Study phase characteristics					
Auditory presentation (vs. visual)	-0.015	0.045	[-0.102, 0.072]	0.15	.70
Bimodal presentation	0.098	0.086	[-0.071, 0.266]	1.37	.24
Production	-0.098	0.086	[-0.267, 0.070]	1.49	.23
Intentional encoding	0.008	0.032	[-0.054, 0.070]	0.08	.78
Self-paced study	-0.018	0.049	[-0.113, 0.078]	0.16	.69
Self-paced study x intentional encoding	0.100	0.076	[-0.048, 0.249]	1.84	.18
Study time per presentation (s)	-0.005	0.005	[-0.014, 0.005]	0.78	.38
Deep orienting task	-0.042	0.030	[-0.101, 0.018]	2.09	.15
Shallow orienting task	0.022	0.038	[-0.053, 0.097]	0.31	.58
Generation task	0.008	0.057	[-0.103, 0.120]	0.02	.88
Divided attention	-0.238	0.111	[-0.456, -0.021]	4.89	.03
Number of memoranda	0.001	< 0.001	[0.001, 0.001]	5.24	.02
Multiple study opportunities	0.010	0.034	[-0.056, 0.077]	0.04	.84
Test phase characteristics					
Intervening cued recall	0.028	0.082	[-0.132, 0.189]	0.11	.74
Intervening free recall	-0.044	0.047	[-0.137, 0.049]	0.75	.39
Continuous recognition	-0.137	0.075	[-0.284, 0.009]	3.71	.05
Retention interval	0.001	0.004	[-0.006, 0.009]	0.09	.77
Continuous recognition x retention interval	-0.006	0.019	[-0.043, 0.031]	0.07	.79
Number of test probes	> -0.001	< 0.001	[-0.001, 0.000]	3.38	.07
Conjunction lures	0.078	0.028	[0.024, 0.132]	8.23	< .01
Component lures	0.128	0.041	[0.047, 0.209]	9.74	< .01
Semantically related lures	0.104	0.023	[0.058, 0.149]	20.26	< .001
Featurally related lures	-0.035	0.028	[-0.091, 0.020]	1.72	.19
Proportion of lures	-0.136	0.102	[-0.335, 0.064]	1.39	.24

Note. SE = standard error. Positive parameter estimates indicate larger age differences (i.e., older adults set especially liberal criteria relative to young adults); negative parameter estimates indicate age differences smaller than the mean (i.e., older adults were relatively less liberal).

Overall performance. Both age groups had a mean recognition criterion (*c*) greater than 0,

indicating a conservative bias to call probes *unstudied/new*. However, this bias was slightly but reliably

weaker in older adults. The model estimated the mean age difference in criterion placement was 0.05 (95% CI: [.02, .07]), $t = 3.13$, $p = .001$. Older adults set a slightly lower, more liberal criterion for terming an item *studied/old* than did young adults, equivalent to an increase of 2% in the hit and false alarm rates given young adults' average criterion ($c = 0.16$).

Further, the slope relating age differences in criterion placement to younger adults' criterion differed significantly from 0. This slope was approximately 0.13 (95% CI: [.06, .19]), indicating that as young adults shifted towards a higher, more conservative criterion, age differences in criterion placement became larger. That is, for a 1 *d'* unit shift towards a more conservative (higher) criterion by younger adults, older adults made a slightly less conservative shift of 0.87 *d'* units.

These relations are depicted in Figure 3, which plots older adults' group-level *c* scores as a function of the corresponding younger adult score. The dotted lines indicate an unbiased criterion placement of 0. The majority of young adult *c* scores lie to the right of the vertical dotted line, indicating a conservative criterion; similarly, the majority of older adult *c* scores lie above the horizontal dotted line. The solid diagonal line represents equivalent criterion placement across age groups. As can be seen, group differences in *c* were less consistent than those in *d'*; although the majority of points fall below the line (indicating older adults had a more liberal criterion than younger adults in that condition), there are also many points above the line (indicating the reverse).
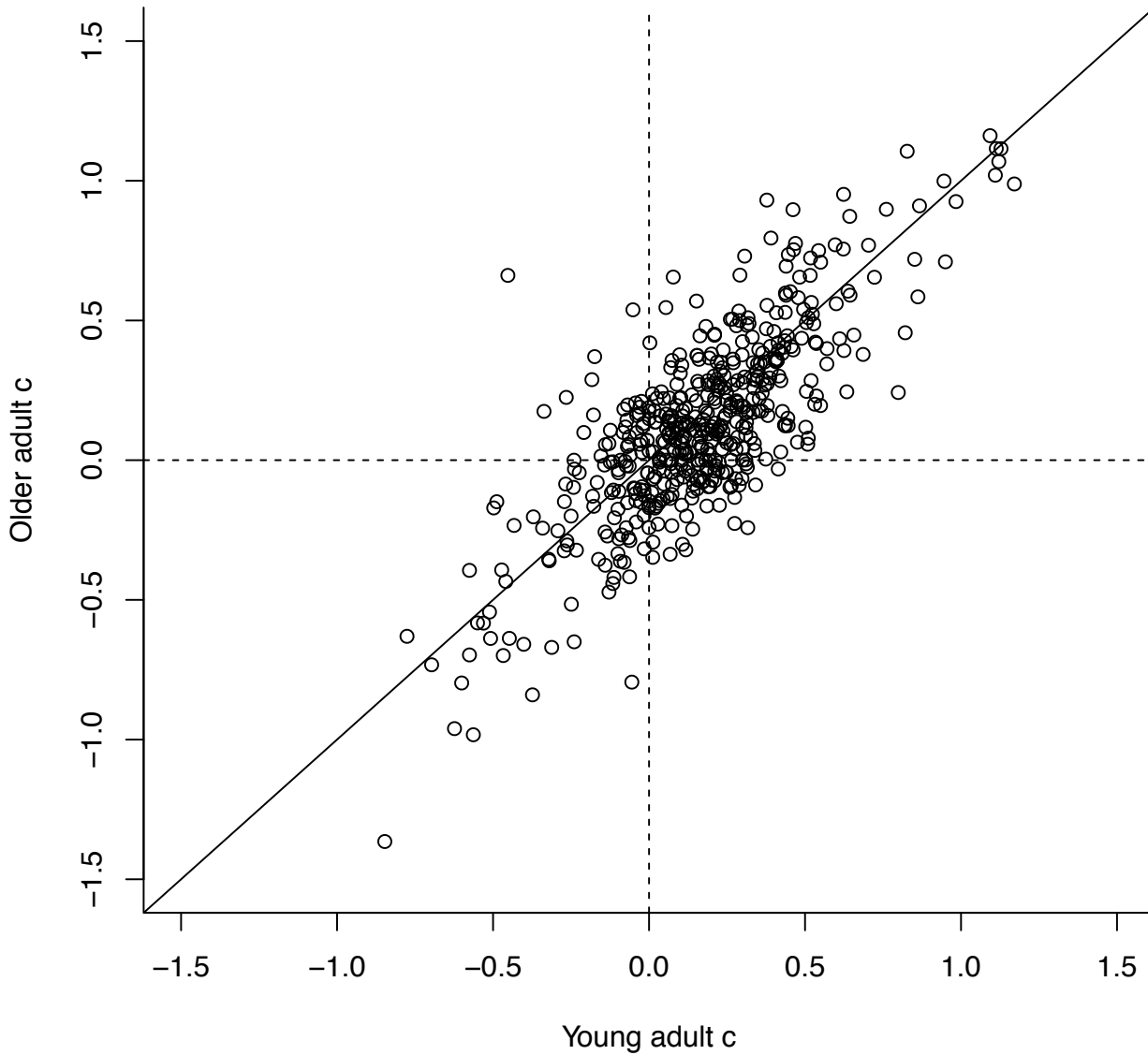
Figure 3. Group-level *c* scores for older adults as a function of group-level *c* scores for younger adults. Each point represents one experimental condition. The diagonal line (identity) represents the relation that would be obtained if there were no age differences in criterion performance. Dotted lines separate positive, conservative criterion values (tendency to term probes *old*) from negative, liberal criterion values (tendency to term probes *new*).

Task characteristics. As with discrimination, a comparison of a fixed-effect versus a random-

effects model revealed there was significant heterogeneity in age differences across conditions[13], $\chi^2_{(1)} =$ 182.0, $p < .001$. Some stimulus and task characteristics were associated with larger age differences in criterion placement than those predicted from the linear relation described above[14].

In particular, lure type robustly influenced age differences in criterion placement, just as it did age differences in discrimination. Compared to younger adults, older adults set significantly lower (i.e., more liberal) criteria for semantically-related lures, conjunction lures, and component lures. The combination of older adults' more liberal response criterion for these lures coupled with the poorer discrimination described above implies that older adults were especially apt to false alarm to these lure types (a conclusion confirmed by the analysis of false alarms in Appendix A). Once again, however, age differences for featurally-related lures did not significantly differ from those for unrelated lures.

Aside from lure type, however, the task characteristics that engendered larger age differences in criterion placement were largely different than those that engendered larger age differences in discrimination, a point we discuss in greater detail in the Discussion. Specifically, compared to young adults, older adults had relatively more liberal criteria for faces and for pictures. They also had more liberal criteria for longer study lists.

However, relative to young adults, older adults set a relatively more conservative criterion for items studied under divided-attention conditions (-0.24 c units more conservative than the baseline age difference). Note that the magnitude of the negative parameter estimate for this effect was greater than the magnitude of the overall positive age difference in criterion placement (mean age difference in c: .05); that is, divided attention conditions actually reversed the overall age difference such that older

---

[13] There was also significant heterogeneity across publications, laboratories, and subject groups. The standard deviation across experiments was 0.04 c units, 0.12 across publications, 0.04 across laboratories, and 0.07 across subject groups within an experiment.

[14] For c, if we excluded the 1% of data with standardized residuals greater than 3, two additional significant effects emerged: (a) the overall age difference was also reversed for continuous-recognition tasks such that older adults were more conservative than young adults in such tasks and (b) older adults set more conservative criteria for longer test lists.

adults were more conservative than younger adults.

The above results suggest that age differences in discrimination are affected by different factors than age differences in criterion placement; for instance, retention interval affects discrimination but not criterion placement. This evidence is indirect insofar as it relies on some variables having an effect on one measure but not another. As an additional test, we added the age difference in d' to the model of c presented above in Table 4. All of the significant influences on age differences in c remained significant even controlling for d', and the age difference in d' itself in a given condition was not a significant predictor of the age difference in c in that same condition ($X^2_{(1)} = 0.76$, $p = .38$).

Unlike with discrimination, we did not find evidence of publication bias; a study's sample size did not predict the observed age difference in criterion placement, $X^2_{(1)} = 0.37$, $p = .54$, and the estimated age difference in criterion placement was in fact larger (at 0.06 c units) in the model that controlled for publication bias. The absence of publication bias in this case is not unexpected given that most studies did not even perform direct comparisons of young and older adults in their criterion placement; thus, there is no reason significant outcomes of this comparison would be favored over non-significant ones.

Lastly, we again examined whether any of the effects above differed for faces in particular. In addition to the overall more liberal criterion that older adults set for faces, three other significant effects emerged. First, for faces only, age differences in criterion were enhanced (i.e., older adults became even more liberal) as per-item study time increased. (Study time did not affect age differences in response criterion for other stimuli.) Second, there was a complex interaction of stimulus type, intentionality of encoding, and self-pacing such that older adults' liberal tendency to respond *old* was greatly enhanced for self-paced, intentional study of faces. Lastly, another complex interaction of stimulus type, retention interval, and test type indicated that longer retention intervals were associated with more conservative responding by older adults, but only for faces studied in continuous recognition

tasks. We are reluctant to over-interpret these results because (a) it is not clear they are predicted a priori by any theory and (b) the large number of such interactions tested—twenty—means that is likely that some would emerge as significant simply by chance.

Discussion

We analyzed 232 prior experiments, totaling over 17,000 research participants, for which we could assess differences between young and older adults in item recognition memory. We examined age differences in their ability to discriminate old (studied) items from new (unstudied) items as well as whether the age groups differed in their overall preference or criterion for judging a particular stimulus as *old*. We found that older adults' recognition memory was generally poorer than young adults; they were consistently less effective at discriminating new and old items. In addition, older adults set a more liberal recognition memory criterion and were overall more apt than young adults to judge items as *old*. Further, each of these age differences was modulated by characteristics of the materials, the task at encoding, and the task at retrieval in ways that inform theories of memory aging. We discuss each of these patterns below.

Age Differences Exist in Recognition Memory

Older adults are less accurate recognizers than younger adults. Across the 232 experiments, which varied in retention interval from seconds to weeks, in stimulus type from faces to words to pictures to texts, in affective valence from positive to neutral to negative, and so on, there was a significant overall main effect of age on recognition memory. Older adults were, on average, 0.46 *d'* units worse than young adults at discriminating old, studied items from new, unstudied items.

This finding is noteworthy because the status of age differences in general recognition memory has in fact been unclear (Craik & McDowd, 1987; Hoyer & Verhaeghen, 2006; Schonfield & Robertson, 1966; Verhaeghen et al., 1993). It has been clear that older adults underperform young adults in recall tasks, perhaps because recall tasks require a large degree of self-initiated processing

(Craik, 1983, 1986; Craik & McDowd, 1987; Danckert & Craik, 2013). Past meta-analyses have also made it evident that older adults are impaired relative to young adults in tasks that require remembering the source or context of information (Spencer & Raz, 1995; Old & Naveh-Benjamin, 2008). However, we could not say with confidence before now that simple item recognition memory declines with age.

The fact that we do observe such a deficit here has important theoretical implications. Specifically, it suggests that memory aging involves, at minimum, a general process that operates across modalities, encoding tasks, stimulus types, and test types. (As we discuss below, memory aging may additionally involve other changes that affect some tasks and materials more than others.) For example, Salthouse (1996) proposed that such changes might arise because of more general slowing in processing speed within and across neural systems. Here, we do not make any specific claims about the source of this global deficit, which the present study was not designed to address; however, we suggest that a general decline in item recognition memory should be accounted for by a successful theory of memory aging.

One starting point for understanding this deficit is the reminder that all tests of recognition are, in fact, tests of context memory to some degree or another. The words and pictures typically used in such studies are rarely wholly novel and are as such recorded against a backdrop of a lifetime of experiences with similar stimuli. Recognizing is in truth localizing a multiply experienced stimulus to the time, place, and circumstances of the relevant study episode. From this perspective, it is hardly surprising that older adults exhibit a deficit, since their longer lives provide a richer and consequently more confusing background of experiences against which to localize that single queried event.

Another clue to the source of this global deficit is that the slope relating older adults' discrimination performance to young adults' discrimination performance is less than unity. That is, age differences were larger for relatively easy memory tasks. (As young adults' performance increased on these easier tasks, older adults' performance did not increase quite as much.) This pattern would

appear to be inconsistent with representational models, such as DRYAD (Benjamin, 2010), in which age differences should be larger for less robustly represented information: Less robustly represented information has little redundancy in memory, so any general decrement of memory with age causes this information to be forgotten. By contrast, information that is well encoded with multiple, redundant representations can survive some degradation of memory and is retained even in old age. This theory would predict the *opposite* pattern of what we observed: In most cases, easy tasks should have resulted in well encoded representations that are resistant to forgetting and thus comparatively small age differences (though it would in fact depend on what made the task easy—short retention intervals would act very differently from, say, multiple study opportunities). Instead, we observed that age differences were largest for easy tasks. This pattern suggests that representational redundancy does not account for the general age difference in recognition memory, and some other account may be necessary.

Although older adults had poorer discrimination performance than young adults in the vast majority of conditions included in the meta-analysis, there were also a handful of cases in which older adults performed equivalently or even better than young adults (e.g., Matzen & Benjamin, 2013). It might be asked whether these data points represent special circumstances in which older adults are actually advantaged relative to young adults. However, as noted above, among all of the variables included in the meta-analysis, we found none that was sufficient to systematically reverse older adults' overall deficit in recognition memory. It is possible that these cases of superior recognition by older adults reflect some other variable, not identified here, that allows older adults to attain item recognition performance equivalent or superior to that of young adults. Another plausible possibility, however, is that these data points simply represent sampling error. Even if, in the population, young adults always outperform older adults in item recognition memory given enough subjects and trials, experiments with a finite sample size might yield some cases in which older adults outperform young adults.

Influences on Age Differences and Theories of Memory Aging

Another important conclusion from the present study is that age differences in memory are greater in some circumstances than others. That is, certain task conditions or stimuli led to an even greater age difference than the global deficit above. Many of these conditions are indeed those that are predicted by specific theoretical accounts of memory aging; in this section, we discuss these effects and their theoretical implications.

For instance, one prominent claim about memory aging is that aging more strongly affects a recollective process of reconstructing a specific study episode, but that age has little or no effect on a cognitively and neutrally dissociable process of responding based on general familiarity (Jacoby, 1999; Jennings & Jacoby, 1997; Yonelinas, 1994). This claim has been supported by past meta-analytic evidence that older adults have difficulty remembering associations between items and their sources or between one item and another (Spencer & Raz, 1995; Old & Naveh-Benjamin, 2008). Our meta-analysis further supported these past results: The single most drastic difference between young and older adults in recognition performance, yielding an age difference of 0.33 *d'* units *over and above* the mean age difference in performance, was for conjunction lures that were novel combinations of two familiar parts. For example, a studied face and a studied hat might be re-combined to form a new face-hat pairing that should be rejected because the *pairing* is new (Vakil, Raz, & Levy, 2010). Such conjunction lures cannot be discriminated from genuine studied pairs (e.g., a previously encountered face-hat pairing) on the basis of simple familiarity because in all cases, all of the parts have studied before and are familiar. Rather, rejecting conjunction lures requires reference to the particular episode or context in which the parts were encountered, and older adults were particularly disadvantaged at that task.

Another prominent theoretical claim is that older adults' memory is more reliant upon semantic knowledge than young adults', possibly because older adults have simply acquired more world

knowledge (Castel, 2005; Zacks & Hasher, 2006). This shift can benefit memory when to-be-remembered information is consistent with prior knowledge, but it can also impair older adults' memory when to-be-remembered memoranda must be distinguished from semantically similar items. (For further discussion of how increasing knowledge over the lifespan can be both helpful and deleterious, see Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014.) The meta-analysis supported this claim as well. Older adults' memory deficits were smaller (though not eliminated) for texts (sentences and paragraphs), which are semantically richer than isolated words and would allow older adults to leverage their semantic knowledge to aid in remembering. Conversely, age differences were larger for faces, which, according to some theoretical perspectives, are individuated at least in part based on more holistic, configural information (for review, Tanaka & Simonyi, 2016). Older adults were also particularly disadvantaged at distinguishing previously studied items from lure items that bore a semantic close relation (such as synonyms of studied words or words from the same semantic category, such as fruits or birds), another finding suggesting that semantics are more central to older than to younger adults' recognition.

Some other extant theories of memory aging received less support from the present analysis. In particular, it has been suggested that older adults are less apt than young adults to initiate helpful mnemonic strategies, such as elaborative encoding, on their own (Craik, 1986; Luo & Craik, 2008). Thus, age differences should be larger on tasks that require self-initiated processing and smaller on tasks for which the environment already guides the learner to helpful strategies. However, we found that age differences were *larger* for tasks in which the experiment required learners to adopt deep, semantic encoding (e.g., judging whether a noun referred to an animate object or an inanimate one), which is generally an effective strategy for learning. This result is inconsistent with the self-initiated processing theory because requiring older adults to use the elaborative encoding strategy should have eliminated the burden of initiating such a strategy on their own and allowed them to obtain

performance more similar to that of young adults. Rather, it appears that young adults were more able to seize the benefits of this type of strategy.

Similarly, the results provided mixed support for motivational accounts that suggest that age differences in memory arise in part because of different goals held by older adults relative to young adults (Carstensen et al., 1999; Mather, 2004). Because older adults perceive their time as more limited, they may be more motivated to spend their time experiencing positive emotions and avoiding negative ones, rather than on trying to acquire new information (which may be more likely to be negative; Garcia, Garas, & Schweitzer, 2012). We found the older adults were indeed disadvantaged at remembering negatively-valenced stimuli relative to neutral stimuli—but only for non-face stimuli, with older adults being comparatively *advantaged* in remembering negative faces relative to neutral. Further, there was no significant evidence that older adults displayed a converse *benefit* to remembering positively-valenced stimuli (faces or otherwise).

Finally, two reliable effects emerged without a strong theoretical prediction. First, older adults had especially poor discrimination performance when there was a longer retention interval between study and test. It is not clear that an a priori prediction of this effect was given by any major theory of memory aging, but it suggests that forgetting rates (e.g., Rubin & Wenzel, 1996; Wixted, 1998) may be different for older adults than for young adults. Second, age differences were diminished when a free recall test intervened between the study phase and the recognition phase. The general benefits of testing on potentiating subsequent memory have been well-documented (Roediger & Karpicke, 2006; Rowland, 2014) although age differences in the size of this effect have not been clearly established (Meyer & Logan, 2013; Rabinowitz & Craik, 1986).

How should these disparate effects be integrated? One possibility is that memory aging reflects several independent changes happening at the same time (Healey & Kahana, 2016). As people get older, their semantic knowledge increases, their source-memory or recollective processes become

impaired, *and* they are motivated to avoid negative stimuli, and *each* of these changes contribute separately to producing differences in recognition memory. The other possibility is that all of these effects—the increasing emphasis on semantic information, the difficulty rejecting conjunction lures, the differential forgetting rates, and so forth—are manifestations of a single underlying age-related change. This possibility has the advantage of parsimony; however, it is not clear that any of the existing theories of memory aging that we discussed above can account for *all* of the observed effects (although many of them certainly account for *some* of the effects). Nevertheless, there is the potential that such a theory might be formulated in the future. We suggest that a fruitful direction for future research will be to determine whether a single age-related change is sufficient to account for all differences in recognition memory or whether memory aging can only be fully accounted for by assuming several different, simultaneous changes.

Finally, we note one important clarification regarding the many variables that did *not* have significant effects on age differences in recognition memory: These variables had small or null effects on the size of the age difference in recognition memory, but that does not mean that these variables have *no* influence on recognition memory in general. Rather, they may have affected recognition memory, but did so equally for the two age groups. For example, we found that requiring participants to self-generate the to-be-remembered stimuli (e.g., from a word-stem clue) did not moderate the size of age differences in discrimination accuracy or response criterion placement. However, this does not mean that generation has *no* effect on memory; in fact, past meta-analytic work has suggested that the generation effect is robust (Bertsch, Pesta, Wiscott, & McDaniel, 2007). Rather, the generation effect simply obtained (roughly) equally for the two age groups.

Age Differences in Recognition Memory Criterion

The meta-analysis also indicated smaller, but nevertheless reliable, differences between age groups in recognition memory criterion placement. Interpreting such effects can be tricky because

criteria often differ across conditions of differential discriminability, especially when those differences are across-subjects (Benjamin, 2001; 2007; Benjamin & Bawa, 2004; Rotello & Macmillan, 2006). Nonetheless, we discuss below how there are many instances in these data in which age-related effects on discriminability and criteria do not go hand-in-hand; such effects are particularly interesting and discussed in greater detail.

Specifically, older adults set a more liberal criterion for terming items *old*—that is, they erred on the side of terming an item *old* or *studied*. This pattern would reduce the number of studied items that older adults miss affirming as studied, at the cost of incorrectly endorsing more studied items. The existence and nature of an age difference in recognition memory criterion has been unclear: Although some other studies also found a more liberal criterion (e.g., Huh et al., 2006), others found similar criteria across ages (e.g., Ahmad, Fernandes, & Hockley, 2015; Baron & Surdy, 1990; Gordon & Clark, 1974) or even a more conservative criterion among older adults (Cassidy & Gutchess, 2015; Criss et al., 2014; Olfman, Light, Schmalstig, Pospisil, Pendergrass, & Chung, 2017). We cannot definitively account for these differences, but we do note that the liberal shift we observe in criterion placement, although statistically reliable in a large meta-analysis, is of small magnitude (0.05 *d'* units). If this small difference is the true population effect size, it is not surprising that some individual studies would observe no difference or even a reversed difference.

Nevertheless, our results are broadly consistent with the claim that older adults are able to flexibly shift their response criterion in ways similar to young adults (e.g., Cassidy & Gutchess, 2015; Criss et al., 2014; Konkel, Selmeczy, & Dobbins, 2015; Olfman et al., 2017; Pendergrass, Olfman, Schmalstig, Seder, & Light, 2012). Older adults generally set more conservative criteria in the same tasks that young adults did, though their criterion shifts were of slightly smaller magnitude.

The above findings characterize older adults' response criterion in general, but criterion placement further varied as a function of the materials, study task, and test task. The age-related liberal

shift in criterion was especially prominent for faces and pictures, for longer lists of to-be-remembered items, and for lists in which the lures were semantically related to the studied items or contained parts of the studied items, but it was diminished (and, in fact, reversed) for divided-attention conditions. These results are striking for several reasons.

First, considering *why* older adults' response criterion might differ in these ways can inform theories of how and why memory changes with age. One possibility is that these apparent criterion shifts are not actually changes in decision strategy per se but reflect changes in processing of the lures (Stretch & Wixted, 2000). For example, in dual-process theories of recognition (e.g., Yonelinas, 1994), recollection of specific episodes should only occur to material that has genuinely been encountered before; however, lures that are insufficiently distinct from targets may erroneously trigger recollection (e.g., Gallo, Foster, & Johnson, 2009), and older adults may be particularly vulnerable to this misrecollection phenomenon (Dodson, Bawa, & Krueger, 2007). Further, an inability to recollect the *true* studied material—e.g., the original stimuli from which a conjunction lure was formed—may also lead to elevated false alarm rates (e.g., Gallo, Bell, Beier, & Schacter, 2006; Pierce, Waring, Schacter, Budson, 2008). These possibilities could plausibly explain older adults' poor performance on tasks involving conjunction lures formed of re-paired elements of the targets and/or lures that bore semantic relations to the targets; age deficits in these tasks were driven mostly (in the case of conjunction lures) or entirely (in the case of semantic lures and component lures) by elevated false alarm rates. That is, even if older adults did not intend to shift their decision strategy, they may have erroneously "recollected" seeing many of the related lures during study (e.g., falsely recollecting *sleep* after studying *rest* and *bed*; Benjamin, 2001; Deese, 1959; Roediger & McDermott, 1995) and/or they may have failed to recollect the original targets (e.g., *blackmail* and *jailbird*) that would help them related reject lures (e.g., the conjunction lure *blackbird* or the semantic lure *criminal*; Matzen & Benjamin, 2013; Matzen, Taylor, & Benjamin, 2011), both of which would result specifically in an elevated false

alarm rate.

However, other apparent age differences in criterion placement were not driven solely by false alarms. Older adults had both higher hit rates and higher false alarms rates for faces and for pictures, which might reflect genuine changes in decision strategy with age. For example, it has been suggested that differences in memory aging in part reflect stereotype threat: Older adults are stereotyped as worse at remembering than young adults, and older adults are concerned about forgetting (Hertzog, 2002; Lineweaver & Hertzog, 1998). If older adults are particularly concerned about forgetting known information (moreso than erroneously "recognizing" an unstudied item), they may reduce such errors by setting a more liberal response criterion. That is, erring on the side of judging a picture as *old* or *studied* is a rational way to avoid failing to recognize a picture that one has indeed seen before. The especially liberal criteria that older adults set for faces may also stem from these concerns, since failing to recognize a known face may be a particularly embarrassing or socially discouraged error. It has also been suggested that inhibitory processes may particularly decline with age (Hasher, Zacks, & May, 1999; Healey, Campbell, & Hasher, 2008; but see Rey-Mermet & Gade, 2017). Although this account is perhaps most directly applicable to intrusions in recall and its application to recognition memory is somewhat more speculative, a decline in inhibitory control may result in older adults erroneously affirming more unstudied lures[15] (Huh et al., 2006), and, indeed, we saw that age differences were larger in false alarm rates than hit rates (see Appendix A). Future work could clarify (e.g., through manipulations of motivation or framing; Rahhal et al., 2001) the degree to which age-related changes in recognition criterion placement reflect changes in motivation, decreased inhibition, or both.

More broadly speaking, the results further support a distinction between the actual fidelity of

---

[15] This interpretation only holds to the extent that one views a recognition memory task as a search for evidence that a stimulus is an old, studied item. If instead the recognition memory task is viewed as one of detecting new items, then poor inhibition might result in a disproportionate rate of new responses.

memory (i.e., the ability to discriminate old versus new probes, discussed above) and the surrounding strategic decisions involved in producing a mnemonic judgment. Indeed, the variables that significantly influenced the magnitude of age differences in response criterion (e.g., divided attention and list length) were substantially different from those that influenced age differences in discrimination (e.g., deep semantic processing, retention interval, and intervening recall tests), with only stimulus type and lure type affecting both measures. Further, these differences across conditions in age differences in response criterion remained intact even when controlling for the age difference in discrimination performance in those same conditions. These patterns suggest that differences in criterion placement arise from at least partially different mechanisms from differences in discrimination and consequently are affected by different variables.

It is less clear what might account for some of the other age differences in criterion placement, such as the fact that older adults' recognition judgments grew especially liberal relative to young adults as the number of to-be-remembered stimuli increased. Although a general effect of list length on memory *accuracy* has been well documented (Strong, 1912, but see Dennis, Lee, & Kinnell, 2008), it is less clear why this variable might affect criterion placement. Thus, a final conclusion is that there is a need for more detailed theories of how and why recognition memory criterion changes with age, including whether these shifts reflect a genuine change in decision strategy versus some other process, such as misrecollection. To date, most theorizing in memory aging has focused on changes in memory fidelity itself. Although no doubt important, memory is always more than just remembering (Benjamin, 2007): The strength of memory itself invariably interacts with decisions about when to study, how to study, how long to study, what cues to submit to memory, and what types of errors are particularly important to avoid (Finley, Tullis, & Benjamin, 2010). To understand how older adults recognize a familiar face, judge whether a newspaper article contains new or known information, or bring to mind their to-do list for the day, there is a need to understand both the fidelity of their memory

itself and the strategic decisions surrounding study and retrieval.

Recognition and Recall in Older Adults

We opened this meta-analysis by noting that age deficits are already well documented in recall tasks (e.g., Craik & McDowd, 1987; Danckert & Craik, 2013); here, we establish that such deficits also exist in recognition. Although the goal of the present meta-analysis was to investigate recognition in and of itself rather than to contrast it with recall, our results nevertheless suggest several implications for the contrast between recall and recognition. First, given that age deficits *do* exist in recognition, it is not necessarily clear that there is a theoretically meaningful division to be drawn between age-related effects on recall and recognition. That is, the core deficit may be one that is general to all memory tasks. Second, to the extent that age deficits *are* larger in recall, it may not be because recall requires self-initiated processing; we did not find strong evidence that older adults performed worse on those tasks requiring more self-initiated processing. Finally, given that older adults *do* show a decline in recognition, we note that recognition—and not just recall—may be useful for testing theories about the underlying mechanisms of age-related deficits. One potential advantage of recognition is that even common recognition tasks permit sensitivity to be distinguished from response criterion to some degree; although it is principle possible to do so in recall as well (e.g., using Type II Signal Detection; Goldsmith & Koriat, 2007; Higham, 2002; Higham & Tam, 2005; Koriat & Goldsmith, 1994), the relevant tasks are used far less widely. Nevertheless, because the present analysis did not constitute a direct comparison between recall and recognition, these conclusions remain somewhat speculative, and there is a need for further analysis—perhaps meta-analysis—of the contrast between recall and recognition in old age.

Theoretical Limitations

Although the large size of the meta-analysis dataset provides important support for the theoretical claims above, the nature of our meta-analysis does place several limitations on

generalization. First, analyzing the published literature meant that we could include and analyze only those variables that were discussed in the original reports, which limited our ability to assess some theories of memory aging. For example, stereotype-threat accounts of cognitive aging (Barber & Mather, 2013; Rahhal et al., 2001) suggest that age differences in recognition memory efficacy should be greater when the task instructions refer to "memory" and "forgetting" and activate a negative stereotype about aging whereas they should be smaller if the task does not reference memory. We could not directly evaluate the effect of stereotype-activation instructions because few Method sections report the exact instructions provided to participants. Similarly, it has been suggested that apparent age decrements in memory might be a function of the time of day at which participants are tested (Hasher et al., 1999; Zacks & Hasher, 2006), but we could not evaluate this hypothesis because very few papers report the time(s) of day at which data was collected (and, indeed, it was likely that this variable was often not systematically manipulated or controlled). We encourage authors of future papers to report the details that might permit a meta-analytic test of these theories. Even among the theories we did assess, there may also be other manipulations or conditions that would provide more effective tests of these theories but that did not appear with sufficient frequency in the literature for meta-analysis. For example, a strong prediction of motivational accounts is that age differences should be modulated by experimentally manipulating the value of particular items (e.g., Castel et al., 2002, 2007; Hargis & Castel, 2018; Hennessee, Knowlton & Castel, 2018; McGillivray & Castel, 2017), but this manipulation did not appear in the literature with sufficient frequency for meta-analysis.

Second, the meta-analysis also did not include every possible interaction among predictor variables, which might have obscured the effects of some potentially relevant variables. For example, we did not observe that presentation modality affected age differences in recognition memory; older adults were neither particularly advantaged nor disadvantaged for materials presented aurally rather than those presented visually. But, it is possible in principle that modality does affect age differences

in recognition memory, but only (for instance) when the study list is particularly long, which would have been captured in a Modality x Number of Memoranda interaction (although we had no reason to predict such an interaction). We did not include every such interaction because the large number of variables included in the meta-analysis would have produced a combinatorial explosion of interactions, inflating the Type I (false positive) error rate and reducing the degrees of freedom, and consequently our statistical power, for detecting the effects that we *were* interested in. Rather, we included only those interactions for which we had specific a priori interest. Nevertheless, it is possible that some other combination of variables might yield effects that were not tested here.

Third, although not an inherent limitation of meta-analysis, this particular analysis concerned only cross-sectional designs in which a group of older adults was compared to a separate group of young adults, rather than longitudinal studies of memory over time. We focused on cross-sectional designs because of their wide use in the study of memory aging, especially for the sort of parametric manipulations of study and test conditions analyzed here, but longitudinal studies (e.g., Hertzog, Dixon, Hultsch, MacDonald, 2002; Park, Lautenschlager, Hedden, Davidson, Smith, & Smith, 2002; Salthouse, 2005, 2016) make their own important contributions to understanding cognitive aging. For instance, although in the present study we did not observe large differences as a function of age within the older-adult group, longitudinal studies may more effectively reveal the specific trajectory of memory change, such as a sharper drop-off at the oldest ages (Salthouse, 2016). Longitudinal studies that trace within-individual change may also more effectively address heterogeneity across individuals, which can be substantial in older-adult populations (and may in fact predict cognitive change in and of itself; MacDonald, Hultsch, & Dixon, 2003).

Lastly, longitudinal studies can address the concern that cross-sectional age comparisons may be influenced by some other, uncontrolled difference across groups; for instance, young adult samples drawn from college populations may be more educated than older adults. In our present meta-analytic

dataset, education does not appear to account for younger adults' superior recognition performance

(although our sample was somewhat more educated than the population as a whole): Among the 153

studies that reported the participants' years of education, the older-adult group was in fact *more*

educated on average ($M = 15.1$ years of education vs. $M = 14.3$ for young adults), and when the

between-group difference in years of education was added into the statistical models, it did not

significantly predict age differences in discrimination, $\chi^2_{(1)} = 0.15$, $p = 69$, nor those in response

criterion, $\chi^2_{(1)} = 1.26$, $p = .26$. Nevertheless, there may be other between-group differences that could

be better accounted for by a longitudinal study. Of course, longitudinal studies have their own

limitations; for instance, repeated testing on similar procedures may allow participants to develop task-

specific strategies (Postman & Schwartz, 1964; Hultsch, 1974). Ultimately, we may best understand

memory aging by combining the strengths and weaknesses of the present analysis with a future meta-

analysis of longitudinal change.

Limitations of Meta-Analysis

Other limitations are intrinsic to meta-analyses in general. In comparing across existing

studies, we did not exert direct experimental control, so our ability to infer causality is limited; it may

be that differences across studies were driven by some other, unidentified variable. Another concern is

the potential existence of publication bias: The data available in the literature may represent a biased

sample of all possible data—for instance, statistically significant results may be more likely to be

published than null or non-significant results—which would correspondingly distort a meta-analysis

based on those data. We conducted Peters' test (J.L. Peters et al., 2006, 2010) and found evidence for a

publication bias favoring significant age differences in discrimination[16], but no such bias for criterion

---

[16] Although this test for publication bias suggests there may be unpublished data with smaller age differences in discrimination, it is unclear where such data, if they exist, may lie: We contacted over 20 prominent labs in memory aging, but we did not find any unpublished data meeting the existing inclusion criteria of the meta-analysis

placement. Nevertheless, even controlling for publication bias, the age difference in discrimination performance and the sources of between-subject heterogeneity remained significant.

One reason that the present dataset may not have been so greatly distorted by a publication bias favoring significant results is that both the presence and absence of age differences are predicted by different theoretical accounts. As noted above, some theoretical accounts propose a global deficit in memory aging (e.g., Salthouse, 1996; Benjamin, 2010) whereas others propose age deficits exist mainly or entirely for specific cognitive processes, such as remembering the source or context of information (e.g, Jennings & Jacoby, 1993; Naveh-Benjamin, 2010; Naveh-Benjamin & Smyth, 2016; Old & Naveh-Benjamin, 2008; Zacks & Hasher, 2006). Thus, both null effects on overall memory (as predicted by a hypothesis of deficits in particular processes) and significant effects on overall memory (as predicted by a global deficit account) would be of interest to the field and likely to be published in the literature; consequently, there is less likely to be a gross publication bias favoring overall differences in d'. Publication bias is even less likely to be a factor in age differences in criterion placement because many manuscripts did not even report tests of age differences in criterion placement; thus, there is no possibility for significant outcomes of this test to be favored over non-significant outcomes. Indeed, if anything, publication bias if anything is likely to be in the direction *against* finding such differences: Researchers are typically more interested in discrimination than criterion placement, and large criterion differences can make it more difficult to observe differences in discrimination (e.g., through floor or ceiling effects), so researchers are likely to avoid experimental designs that yield substantial differences in criterion placement.

However, in line with past evaluations of meta-analysis (McShane et al. 2016), we emphasize that inferences about publication bias and its consequences are only inferences, since the set of unpublished studies is ultimately unknown. Concerns about publication bias can be fully mitigated only when all of the studies or data are known, such as through pre-registered reports.

Methodological Implications

This caveat points to one other important limitation of the present study: We were limited to data that allowed us to obtain measures of both discrimination accuracy and response criterion. Although this included a large number of experiments (232), a number of other experiments were potentially relevant to the meta-analysis but had to be excluded because the authors reported only a single dependent measure, such as only hit rates or only a difference score between hits and false alarms—both of which are insufficient to disentangle discrimination from response criterion. Further, even the studies we did include generally allowed us only to compute d' and c as descriptive indices of discrimination and criterion placement, rather than the more accurate measures that could be obtained from more nuanced unequal-variance (Wixted, 2007) or dual-process (Yonelinas, 1994) models. Fortunately, the experiments with sufficient data for inclusion are likely to be a reasonably representative subset of all experiments; it is unlikely that the studies excluded for insufficient data systematically differed in their experimental design or outcomes. (That is, the studies for which insufficient data was reported were unlikely to systematically differ in, say, the retention interval between study and test.) Thus, the conclusions of the meta-analysis are unlikely to be drastically altered by these omissions. Nevertheless, it would be helpful if researchers more consistently reported measures of both discrimination and criterion (or data that would allow such measures to be computed), and perhaps even those that would allow more detailed models of the decision process to be fit to the data (e.g., by obtaining confidence ratings). The meta-analysis could have included an even larger data set (and hence obtained more precise estimates), and it would mitigate any concerns that the remaining experiments were in some way non-representative of all experiments. Indeed, given the evidence discussed above that age differences in discrimination versus those in response criterion are affected by different variables and may reflect different underlying causes, researchers may find it theoretically informative to test hypotheses about both discrimination and criterion placement.

In addition, for most of the experiments included in the analysis, we could only compare the group-level d' and c scores (e.g., the d' for all older adults in a condition versus the d' for all younger adults in a condition) because in very few cases did we have data on how much those scores varied within a group. That is, it would be possible to compute a d' score for each individual participant and then assess the standard deviation of these scores across each of the individuals in the older adult group (and, analogously, across each individual in the young adult group), thus avoiding the potential hazards of working with group-level data (Estes & Maddox, 2005).

Thus, one conclusion is that small additions to the data reported in a typical recognition memory paper could substantially enhance future meta-analyses as well as the analyses contained within the papers themselves. We suggest that it would be helpful for future analyses of recognition memory to include both measures of memory fidelity and criterion placement—perhaps those obtained from sophisticated models of the recognition decision process—and to include data on how these measures vary across individuals at the subject level. In addition, as noted above, given claims that age differences in recognition memory may be driven by the instructions (Rahhal et al., 2001) or by the time of day (Hasher et al., 1999), it would also aid theoretical advancement to discuss these variables more routinely in Method sections.

Conclusions

One of the most basic and fundamental tasks in memory is recognition: deciding whether a particular face, word, text, or picture has been encountered before. This ability is not constant over the lifespan. A meta-analysis of item recognition memory experiments indicated important differences in how young and older adults judge items as old or new. Recognition memory generally declined in older adults relative to young adults, resulting in a reduced ability to discriminate new items from old items. These declines were particularly acute for recognition tasks including lures that were comprised of familiar, studied parts and for lures semantically related to the targets, but they were smaller for

semantically rich materials. There was also some evidence that older adults were particularly

disadvantaged at remembering negatively-valenced stimuli (except faces). Together, these results

support several existing theoretical claims about memory aging: Older adults have particular difficulty

remembering contextual or associative information, older adults rely more heavily on semantic

information, and older adults are motivated to not attend to negative stimuli. In addition, older adults

tended to set more liberal recognition criteria that led them to err on the side of terming items *old* rather

than *new*, which might in part reflect concerns about increased forgetting with age. More broadly,

these data represent a comprehensive look at the factors that constrain the aging of item recognition

memory and constrain future development of theories about the cause or causes of memory aging.

References

References marked with an asterisk indicate studies included in the meta-analysis

*Ahmad, F. N., Fernandes, M., & Hockley, W. E. (2015). Improving associative memory in older

adults with unitization. *Aging, Neuropsychology, and Cognition*, *22*(4), 452-472. doi:

10.1080/13825585.2014.980216

*Algarabel, S., Escudero, J., Mazón, J. F., Pitarque, A., Fuentes, M., Peset, V., & Lacruz, L. (2009).

Familiarity-based recognition in the young, healthy elderly, mild cognitive impaired and

Alzheimer's patients. *Neuropsychologia*, *47*, 2056–2064.

doi:10.1016/j.neuropsychologia.2009.03.016

*Anastasi, J. S., & Rhodes, M. G. (2006). Evidence for an own-age bias in face recognition. *North

American Journal of Psychology*, *8*, 237–252. doi:10.3758/BF03206441

*Anderson, B. A., Jacoby, L. L., Thomas, R. C., & Balota, D. A. (2011). The effects of age and divided

attention on spontaneous recognition. *Memory & Cognition*, *39*, 725–735. doi:10.3758/s13421-

010-0046-z

*Angel, L., Bastin, C., Genon, S., Balteau, E., Phillips, C., Luxen, A., … Collette, F. (2013).

Differential effects of aging on the neural correlates of recollection and familiarity. *Cortex*, *49*(6),

1585–1597. doi:10.1016/j.cortex.2012.10.002

*Arvind Pala, P., N'Kaoua, B., Mazaux, J. M., Simion, A., Lozes, S., Sorita, E., & Sauzéon, H. (2014).

Everyday-like memory and its cognitive correlates in healthy older adults and in young patients

with traumatic brain injury: A pilot study based on virtual reality. *Disability and Rehabilitation:

Assistive Technology*, *9*(6), 463–473. doi:10.3109/17483107.2014.941952

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK:

Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random

effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

*Bäckman, L. (1991). Recognition memory across the adult life span: the role of prior knowledge. *Memory & Cognition*, *19*, 63–71. doi:10.3758/BF03198496

*Badham, S. P., Hay, M., Foxon, N., Kaur, K., & Maylor, E. A. (2016). When does prior knowledge disproportionately benefit older adults' memory? *Aging, Neuropsychology, and Cognition*, *23*, 338-365. doi: 10.1080/13825585.2015.1099607

*Badham, S. P., & Maylor, E. A. (2011). Age-related associative deficits are absent with nonwords. *Psychology and Aging*, 26, 689-694. doi:10.1037/a0022205

*Ballesteros, S., & Mayas, J. (2015). Selective attention affects conceptual object priming and recognition: A study with young and older adults. *Frontiers in Psychology*, *6*(JAN), 1–11. doi:10.3389/fpsyg.2015.00567

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*, 81.

Barber, S. J., & Mather, M. (2013). Stereotype threat can enhance, as well as impair, older adults' memory. *Psychological Science*, *24*, 2522-2529. doi: 10.1177/0956797613497023

Baron, A., & Surdy, T. M. (1990). Recognition memory in older adults: Adjustment to changing contingencies. *Journal of the Experimental Analysis of Behavior*, *54*(3), 201-212. doi:10.1901/jeab.1990.54-201

*Bartlett, J. C., Leslie, J. E., Tubbs, A., & Fulton, A. (1989). Aging and memory for pictures of faces. *Psychology and Aging*, *4*, 276–283. doi:10.3758/BF03197012

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

*Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: a study of the effects of test format and aging. *Neuropsychology*, *17*, 14–24. doi:10.1037/0894-4105.17.1.14

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using

    lme4. *Journal of Statistical Software*, *67,* 1-48. doi: 10.18637/jss.v067.i01

Benjamin, A. S. (2007). Memory is more than just remembering: Strategic control of encoding,

    accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The*

    Psychology of Learning and Motivation: Vol. 48. Skill and Strategy in Memory Use (pp.175-223).

    London: Academic Press. doi: 10.106/S0079-7421(07)48005-7

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The

    DRYAD theory of aging and memory judgments. *Psychological Review*, *117*, 1055-1079. doi:

    10.1037/a0020810

Benjamin, A. S. (2016). Aging and associative recognition: A view from the DRYAD model of age-

    related memory deficits. *Psychology and Aging*, *31*, 14-20. doi: 10.1037/pag0000065

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition.

    Journal of Memory and Language, *51*, 159-172.

*Benjamin, A. S., & Craik, F. I. M. (2001). Parallel effects of aging and time pressure on memory for

    source: evidence from the spacing effect. *Memory & Cognition*, *29*, 691–697.

    doi:10.3758/BF03200471

Benjamin, A. S., Diaz, M., Matzen, L. E., & Johnson, B. (2012). Tests of the DRYAD theory of the

    age-related deficit in memory for context: Not about context, and not about aging. *Psychology*

    and Aging, *27*, 418-428. doi: 10.1037/a0024786

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to

    recognition memory. *Psychological Review*, *116*, 84.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic

    review. *Memory & Cognition*, *35*, 201-210.

*Berry, J. M., Williams, H. L., Usubalieva, A., & Kilb, A. (2013). Metacognitive awareness of the

associative deficit for words and names. *Aging, Neuropsychology and Cognition*, *20*(5), 592–619. doi:10.1080/13825585.2012.761670

*Bowman, C. R., & Dennis, N. A. (2015). Age differences in the neural correlates of novelty processing: The effects of item-relatedness. *Brain Research*, 1612, 2-15. doi: 10.1016/j.brainres.2014.08.006

*Bryce, M. S., & Dodson, C. S. (2013). Cross-age effect in recognition performance and memory monitoring for faces. *Psychology and Aging*, *28*, 87-98. doi:10.1037/a0030213

*Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition*, *51*, 251–261. doi:10.1016/S0278-2626(03)00030-7

*Budson, A. E., Wolk, D. A., Chong, H., & Waring, J. D. (2006). Episodic memory in Alzheimer's disease: Separating response bias from discrimination. *Neuropsychologia*, *44*, 2222–2232. doi:10.1016/j.neuropsychologia.2006.05.024

*Bugaiska, A., Clarys, D., Jarry, C., Taconnat, L., Tapia, G., Vanneste, S., & Isingrini, M. (2007). The effect of aging in recollective experience: The processing speed and executive functioning hypothesis. *Consciousness and Cognition*, *16*, 797–808. doi:10.1016/j.concog.2006.11.007

*Bunce, D. (2003). Cognitive support at encoding attenuates age differences in recollective experience among adults of lower frontal lobe function. *Neuropsychology*, *17*, 353–361. doi:10.1037/0894-4105.17.3.353

*Bunce, D., & Macready, A. (2005). Processing speed, executive function, and age differences in remembering and knowing. *Quarterly Journal of Experimental Psychology*, *58*, 155–168.

*Carr, V. A., Castel, A. D., & Knowlton, B. J. (2015). Age-related differences in memory after attending to distinctiveness or similarity during learning. *Aging, Neuropsychology, and Cognition*, *22*(2), 155–169. doi:10.1080/13825585.2014.898735

*Carson, N., Murphy, K. J., Moscovitch, M., & Rosenbaum, R. S. (2016). Older adults show a self-reference effect for narrative information. *Memory*, *24*, 1157–1172. doi:10.1080/09658211.2015.1080277

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist*, *54*, 165-181. doi: 10.1037/0003-066X.54.3.165

Carter, E.C., Schönbrodt, F.D., Hilgard, J., & Gervais, W.M. (2017). Correcting for bias in psychology: A comparison of meta-analysis methods. Manuscript under review.

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: the role of schematic support. *Psychology and Aging*, *20*, 718.

Castel, A. D. (2007). Aging and memory for numerical information: The role of specificity and expertise in associative memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*, P194-P196.

Castel, A. D., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078-1085.

*Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging*, *18*, 873-885. doi: 10.1037/0882-7974.18.4.873

Castel, A. D., McGillivray, S., & Worden, K. M. (2013). Back to the future: Past and future era-based schematic support and associative memory for prices in younger and older adults. *Psychology and Aging*, *28*, 996-1003. doi: 10.1037/a0034160

*Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*, 403–416. doi:10.3758/BF03200930

*Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable

nature of negative images for older adults. *Journal of Experimental Psychology: General*, *132*, 310–324. doi:10.1037/0096-3445.132.2.310

*Charlot, V., & Feyereisen, P. (2004). Aging and the deletion of function of inhibiton. *Aging, Neuropsychology, and Cognition*, *11*, 12–24.

*Chasteen, A. L., Bhattacharyya, S., Hortota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. *Experimental Aging Research*, *31*, 235–260. doi:10.1017/CBO9781107415324.004

*Chung, C., & Light, L. L. (2009). Effects of age and study repetition on plurality discrimination. *Aging, Neuropsychology and Cognition*, *16*, 446–460. doi:10.1080/13825580902773875

*Clarys, D., Bugaiska, A., Tapia, G., & Baudouin, A. (2009). Ageing, remembering, and executive function. *Memory*, *17*, 158–168. doi:10.1080/09658210802188301

*Coane, J. H., Huff, M. J., & Hutchison, K. A. (2016). The ironic effect of guessing: Increased false memory for mediated lists in younger and older adults. *Aging, Neuropsychology, and Cognition*, *23*(3), 282–303. doi:10.1080/13825585.2015.1088506

*Coane, J. H., Sánchez-Gutiérrez, C., Stillman, C. M., & Corriveau, J. A. (2014). False memory for idiomatic expressions in younger and older adults: Evidence for indirect activation of figurative meanings. *Frontiers in Psychology*, *5*(JUL), 1–9. doi:10.3389/fpsyg.2014.00764

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, *15*, 692-712.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Cohn, M., Emrich, S. M., & Moscovitch, M. (2008). Age-related deficits in associative memory: The influence of impaired strategic retrieval. *Psychology and Aging*, *23*, 93–103. doi:10.1037/0882-7974.23.1.93

*Collette, F., Grandjean, J., Lorant, C., & Bastin, C. (2014). The role of nemory traces quality in

    directed forgetting: A comparison of young and older participants. *Psychologica Belgica*, *54*(4),

    310–327. doi:10.5334/pb.au

*Comblain, C., D'Argembeau, A., Van der Linden, M., & Aldenhoff, L. (2004). The effect of ageing on

    the recollection of emotional and neutral pictures. *Memory*, *12*, 673–684.

    doi:10.1080/09658210344000477

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage

    Foundation.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory.

    Philosophical Transactons of the Royal Society of London, *Series B, 302*, 341-359.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Flix & H. Hagendorf

    (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 409-422).

    Amsterdam: North-Holland.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of

    Experimental Psychology: Learning, Memory, and Cognition*, *13*, 474-479.

*Craik, F. I. M., & Schloerscheidt, A. M. (2011). Age-related differences in recognition memory:

    Effects of materials and context change. *Psychology and Aging*, 26, 671-677.

    doi:10.1037/a0022203

*Criss, A. H., Aue, W., & Kılıç, A. (2014). Age and response bias: Evidence from the strength-based

    mirror effect. *The Quarterly Journal of Experimental Psychology*, 67(10), 1910-1924. doi:

    10.1080/17470218.2013.874037

*Cyr, A.-A., & Anderson, N. D. (2012). Trial-and-error learning improves source memory among

    young and older adults. *Psychology and Aging*, *27*, 429–439. doi:10.1037/a0025115

*Danckert, S. L., & Craik, F. I. M. (2013). Does aging affect recall more than recognition memory?

Psychology and Aging, *28*(4), 902–909. doi:10.1037/a0033263

*Daniels, K. A., Toth, J. P., & Hertzog, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology and Aging*, *24*, 494–500. doi:10.1037/a0015269

*D'Argembeau, A., & Van der Linden, M. (2004). Identity but not expression memory for unfamiliar faces is affected by ageing. *Memory*, *12*, 644–654. doi:10.1017/CBO9781107415324.004

*Deason, R. G., Hussey, E. P., Ally, B. A., & Budson, A. E. (2012). Changes in response bias with different study-test delays: evidence from young adults, older adults, and patients with Alzheimer's disease. *Neuropsychology*, *26*, 119–126. doi:10.1037/a0026330

*Denney, N. W., Dew, J. R., & Kihlstrom, J. F. (1992). An adult developmental study of the encoding of spatial location. *Experimental Aging Research*, *18*, 25–32. doi:10.1080/03610739208253907

*Dennis, N. A., Hayes, S. M., Prince, S. E., Madden, D. J., Huettel, S. A., & Cabeza, R. (2008). Effects of aging on the neural correlates of successful item and source memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 791–808. doi:10.1037/0278-7393.34.4.791

*Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia*, *45*, 3157-3166. doi:10.1016/j.neuropsychologia.2007.07.003

Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, *59*, 361-376. doi: j.jml.2008.06.007

*Dew, I. T. Z., & Giovanello, K. S. (2010). Differential age effects for implicit and explicit conceptual associative memory. *Psychology and Aging*, *25*, 911-921. doi:10.1037/a0019940

*Duarte, A., Graham, K. S., & Henson, R. N. (2010). Age-related changes in neural activity associated with familiarity, recollection and false recognition. *Neurobiology of Aging*, *31*, 1814–1830. doi:10.1016/j.neurobiolaging.2008.09.014

*Duarte, A., Henson, R. N., & Graham, K. S. (2008). The effects of aging on the neural correlates of

subjective and objective recollection. *Cerebral Cortex*, *18*, 2169–2180.

doi:10.1093/cercor/bhm243

*Duverne, S., Motamedinia, S., & Rugg, M. D. (2009a). Effects of age on the neural correlates of

retrieval cue processing are modulated by task demands. *Journal of Cognitive Neuroscience*, *21*,

1–17. doi:10.1162/jocn.2009.21001

*Duverne, S., Motamedinia, S., & Rugg, M. D. (2009b). The relationship between aging, performance,

and the neural correlates of successful memory encoding. *Cerebral Cortex*, *19*, 733–744.

doi:10.1093/cercor/bhn122

*Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: differences in susceptibility to

false fame. *Psychology and Aging*, *5*, 379–387. doi:10.1037/0882-7974.5.3.379

*Ebner, N. C., Riediger, M., & Lindenberger, U. (2009). Schema reliance for developmental goals

increases from early to late adulthood: improvement for the young, loss prevention for the old.

Psychology and Aging, *24*, 310–23. doi:10.1037/a0015430

Egan, J. P. (1958). *Recognition memory and the operating characteristic*. (Tech. Note No. AFCRC-TN-

58-51, AO-152650). Bloomington, IN: Indiana University Hearing and Communication

Laboratory.

*El Haj, M., Raffard, S., & Gély-Nargeot, M.-C. (2016). Destination memory and cognitive theory of

mind in normal ageing. *Memory*, *24*(4), 526–534. doi:10.1080/09658211.2015.1021257

*Emery, L., & Hess, T. M. (2008). Viewing instructions impact emotional memory differently in older

and young adults. *Psychology and Aging*, *23*, 2–12. doi:10.1037/0882-7974.23.1.2

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from

model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*, 403-

408.

*Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant

cues. *Psychology and Aging*, *7*, 443–452.

*Fernandes, M. A., Pacurar, A., Moscovitch, M., & Grady, C. (2006). Neural correlates of auditory recognition under full and divided attention in younger and older adults. *Neuropsychologia*, *44*, 2452–2464. doi:10.1016/j.neuropsychologia.2006.04.020

*Ferris, S. H., Crook, T., Clark, E., McCarthy, M., & Rae, D. (1980). Facial recognition memory deficits in normal aging and senile dementia. *The Journal of Gerontology*, *35*, 707–714.

*Feyereisen, P. (2009). Enactment effects and integration processes in younger and older adults' memory for actions. *Memory*, *17*, 374–385. doi:10.1080/09658210902731851

Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2010). Metacognitive control of learning and remembering. In M.S. Khine & I. M. Saleh (Eds.) *New science of learning: cognition, computers and collaboration in education* (pp. 109-131). New York, NY: Springer. doi: 10.1007/978-1-4419-5716-0_6

*Fischer, H., Nyberg, L., & Bäckman, L. (2010). Age-related differences in brain regions supporting successful encoding of emotional faces. *Cortex*, *46*, 490–497. doi:10.1016/j.cortex.2009.05.011

*Fleischman, D. A., Gabrieli, J. D. E., Wilson, R. S., Moro, T. T., & Bennett, D. A. (2005). Repetition priming and recognition memory in younger and older persons: temporal stability and performance. *Neuropsychology*, *19*, 750–759. doi:10.1037/0894-4105.19.6.750

*Flicker, C., Ferris, S. H., Crook, T., & Bartus, R. T. (1990). Age differences in the vulnerability of facial recognition memory to proactive interference. *Experimental Aging Research*, *15*, 189-194.

*Freeman, J. E., & Ellis, J. A. (2003). Aging and the accessibility of performed and to-be-performed actions. *Aging, Neuropsychology, and Cognition*, *10*, 298–309.

*Friedman, D., de Chastelaine, M., Nessler, D., & Malcolm, B. (2010). Changes in familiarity and recollection across the lifespan: An ERP perspective. *Brain Research*, *1310*, 124–141. doi:10.1016/j.brainres.2009.11.016

*Friedman, D., & Trott, C. (2000). An event-related potential study of encoding in young and older

adults. *Neuropsychologia*, *38*, 542–557. doi:10.1016/S0028-3932(99)00122-0

*Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based

retrieval monitoring: the importance of event distinctiveness. *Psychology and Aging*, *22*, 209–13.

doi:10.1037/0882-7974.22.1.209

*Gavazzeni, J., Andersson, T., Bäckman, L., Wiens, S., & Fischer, H. (2012). Age, gender, and arousal

in recognition of negative and neutral pictures 1 year later. *Psychology and Aging*, *27*, 1039–

1052. doi:10.1037/a0027946

Garcia, D., Garas, A., & Schweitzer, F. (2012). Positive words carry less information than negative

words. *EPJ Data Science*, *1*, 1-12. doi: 10.1140/epjds3

*Glisky, E. L., & Marquine, M. J. (2009). Semantic and self-referential processing of positive and

negative trait adjectives in older adults. *Memory*, *17*, 144–157. doi:10.1080/09658210802077405

*Gopie, N., Craik, F. I. M., & Hasher, L. (2010). Destination memory impairment in older people.

Psychology and Aging, 25, 922-928. doi:10.1037/a0019703Goldsmith, M., & Koriat, A. (2007).

The strategic regulation of memory accuracy and informativeness. In A. S. Benjamin & B. H.

Ross (Eds.), *The psychology of learning and motivation: Vol. 48. Skill and strategy in memory use*

(pp. 1–60). San Diego, CA: Elsevier. doi: 10.1016/S0079-7421(07)48001-X

*Gordon, S. K., & Clark, W. C. (1974). Adult age differences in word and nonsense syllable

recognition memory and response criterion. *Journal of Gerontology*, 29(6), 659-665. doi:

10.1093/geronj/29.6.659

*Grady, C. L., Hongwanishkul, D., Keightley, M., Lee, W., & Hasher, L. (2007). The effect of age on

memory for emotional faces. *Neuropsychology*, *21*, 371–380. doi:10.1037/0894-4105.21.3.371

*Gras, D., Tardieu, H., Piolino, P., & Nicolas, S. (2011). Presentation modality effect on false

memories in younger and older adults: the use of an inference paradigm. *Memory*, *19*, 92–102.

doi:10.1080/09658211.2010.537278

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

*Gregory, M. E., Mergler, N. L., Durso, F. T., & Zandi, T. (1988). Cognitive reality monitoring in adulthood. *Educational Gerontology*, *14*, 1–13.

*Grühn, D., Scheibe, S., & Baltes, P. B. (2007). Reduced negativity effect in older adults' memory for emotional pictures: the heterogeneity-homogeneity list paradigm. *Psychology and Aging*, *22*, 644–649. doi:10.1037/0882-7974.22.3.644

*Guez, J., & Naveh-Benjamin, M. (2016). Proactive interference and concurrent inhibitory processes do not differentially affect item and associative recognition: Implication for the age-related associative memory deficit. *Memory*, *24*(8), 1091–1107. doi:10.1080/09658211.2015.1069852

*Gutchess, A. H., Kensinger, E. A., Yoon, C., & Schacter, D. L. (2007). Ageing and the self-reference effect in memory. *Memory*, 15(8), 822-837. doi: 10.1080/09658210701701394

*Gutchess, A. H., Kensinger, E. A., & Schacter, D. L. (2007). Aging, self-referencing, and medial prefrontal cortex. *Social Neuroscience*, 2(2), 117-133. doi: 10.1080/17470910701399029

*Hämmerer, D., Hopkins, A., Betts, M. J., Maaß, A., Dolan, R. J., & Düzel, E. (2017). Emotional arousal and recognition memory are differentially reflected in pupil diameter responses during emotional memory for negative events in younger and older adults. *Neurobiology of Aging*, *58*, 129–139. doi:10.1016/j.neurobiolaging.2017.06.021

*Hanaki, R., Abe, N., Fujii, T., Ueno, A., Nishio, Y., Hiraoka, K., … Mori, E. (2011). The effects of aging and Alzheimer's disease on associative recognition memory. *Neurological Sciences*, *32*, 1115–1122. doi:10.1007/s10072-011-0748-4

Hargis, M. B., & Castel, A. D. (2018). Younger and older adults' associative memory for medication interactions of varying severity. *Memory*, *26*, 1151-1158. doi: doi.org/10.1080/09658211.2018.1441423

*Harkins, S. W., Chapman, C. R., & Eisdorfer, C. (1979). Memory loss and response bias in senescence. *The Journal of Gerontology*, *34*, 66–72.

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

*Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging*, *4*, 106–112.

*Hashtroudi, S., Johnson, M. K., Vnek, N., & Ferguson, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology and Aging*, *9*, 160–170.

*He, Y., Ebner, N. C., & Johnson, M. K. (2011). What predicts the own-age bias in face recognition memory? *Social Cognition*, *29*, 97–109.

Healey, M. K., Campbell, K. L., & Hasher, L. (2008). Cognitive aging and increased distractibility: Costs and potential benefits. In W.S. Sossin, J.-C. Lacaille, V.F. Castellucci, & S. Belleville (Eds.), *Progress in brain research* (Vol. 169, pp. 353-363). Amsterdam, The Netherlands: Elsevier. doi: 10.1016/S0079-6123(07)00022-2

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, *123*, 23-69. doi: 10.1037/rev0000015

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39-65. doi: 10.1002/jrsm.5

Hedges, L. V., & Vevea, J. L. (1998). Fixed and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486-504.

Hennessee, J. P., Knowlton, B. J., & Castel, A. D. (2018). The effects of value on context-item associative memory in younger and older adults. *Psychology and Aging*, *33*, 46-56. doi: 10.1037/pag0000202

Hertzog, C. (2002). Metacognition in older adults: implications for application. In T. L. Perfect & B. L.

    Schwartz (Eds.), *Applied metacognition* (pp. 169-196). Cambridge, UK: Cambridge University

    Press.

*Hess, T. M. (1985). Aging and context influences on recognition memory for typical and atypical

    script actions. *Developmental Psychology*, *21*, 1139–1151. doi:10.1037//0012-1649.21.6.1139

*Hess, T. M., & Slaughter, S. J. (1990). Schematic knowledge influences on memory for scene

    information in young and older adults. *Developmental Psychology*, *26*, 855–865.

    doi:10.1037/0012-1649.26.5.855

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the

    encoding specificity principle revisited. *Memory & Cognition*, *30*, 67-80.

Higham, P. A., & Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal*

    of Memory and Language, *52*, 595-617.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory

    model. *Psychological Review*, *95*, 528-551. doi: 10.1037/0033-295X.95.4.528

*Howard, D. V., Heisey, J. G., & Shaw, R. J. (1986). Aging and the priming of newly learned

    associations. *Developmental Psychology*, *22*, 78–85. doi:10.1037//0012-1649.22.1.78

*Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W.J. (2006). Aging selectively impairs

    recollection in recognition memory for pictures: Evidence from modeling and ROC curves.

    Psychology and Aging, *21*, 96-106. doi: 10.1037/0882-7974.21.1.96

Hoyer, W. J., & Verhaeghen, P. (2005). Memory aging. In J. Birren & K. Schaie (Eds.), *Handbook of*

    the psychology of aging (6th ed., pp. 209-232). New York: Elsevier.

Huh, T. J., Kramer, J. H., Gazzaley, A., & Delis, D. C. (2006). Response bias and aging on a

    recognition memory task. *Journal of the International Neuropsychological Society*, *12*, 1-7.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models:

implications for cumulative research knowledge. *International Journal of Selection and Assessment*, *8*, 275-292.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research synthesis* (2nd ed.). Thousand Oaks, CA: Sage.

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124. doi: 10.1371/journal.pmed.0020124

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

Jacoby, L. L. (1999). Ironic effects of repetition: measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 3-22.

*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*, 493–504. doi:10.1016/j.jml.2005.01.007

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: aging, attention, and control. *Psychology and Aging*, *8*, 283-293. doi: 10.1037/0882-7974.8.2.283

*Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: telling effects of repetition. *Psychology and Aging*, *12*, 352–361. doi:10.1037/0882-7974.12.2.352

*Johnson, M. K., De Leonardis, D. M., & Hashtroudi, S. (1995). Aging and single versus multiple cues in source monitoring. *Psychology and Aging*, *10*, 507–517.

*Johnson, M. M., Schmitt, F. A., & Pietrukowicz, M. (1989). The memory advantages of the generation effect: age and process differences. *The Journal of Gerontology*, *44*, P91–P94. doi: 10.1093/geronj/44.3.P941

*Jones, T. C., & Jacoby, L. L. (2005). Conjunction errors in recognition memory: Modality-free errors

for older adults but not for young adults. *Acta Psychologica*, *120*, 55–73.

doi:10.1016/j.actpsy.2005.03.003

*Kapucu, A., Rotello, C. M., Ready, R. E., & Seidl, K. N. (2008). Response bias in "remembering"

emotional stimuli: A new perspective on age differences. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *34*, 703–711. doi:10.1037/0278-7393.34.3.703

*Karlsson, T., Adolfsson, R., Börjesson, A., & Nilsson, L. G. (2003). Primed word-fragment

completion and successive memory test performance in normal aging. *Scandinavian Journal of*

Psychology, *44*, 355–361. doi:10.1111/1467-9450.00355

*Kensinger, E. A. (2008). Age differences in memory for arousing and nonarousing emotional words.

The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, *63B*, P13–

P18. doi:10.1093/geronb/63.1.P13

*Kim, S.-Y., & Giovanello, K. S. (2011). The effects of attention on age-related relational memory

deficits: fMRI evidence from a novel attentional manipulation. *Journal of Cognitive*

Neuroscience, *23*, 3637–3656. doi:10.1162/jocn_a_00058

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distinguishing the

accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of*

Experimental Psychology: General, *123*, 297-315.

*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*, 189–193.

doi:10.1111/1467-9280.01441

*Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review*, *13*, 84–91.

doi:10.3758/BF03193817

*Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False

recognition of abstract versus common objects in older and younger adults: testing the semantic

categorization account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 499–510. doi:10.1037/0278-7393.29.4.499

*Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*, 555–583. doi:10.1006/jmla.1997.2529

*Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: effects of target age and memory load. *Memory & Cognition*, *33*, 1017–1024. doi:10.3758/BF03193209

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

*Larsson, M., & Bäckman, L. (1993). Semantic activation and episodic odor recognition in young and older adults. *Psychology and Aging*, *8*, 582–588. doi:10.1037/0882-7974.8.4.582

*LaVoie, D. J., & Faulkner, K. M. (2008). Production and identification repetition priming in amnestic mild cognitive impairment. Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 15, 523–544. doi:10.1080/13825580802051497

*Lazzara, M. M., Yonelinas, A. P., & Ober, B. A. (2002). Implicit memory in aging: Normal transfer across semantic decisions and stimulus format. *Aging, Neuropsychology, and Cognition*, *9*, 145–156. doi:10.1076/anec.9.2.145.9545

*Lee, Y.-S., Lee, C.-L., & Yang, H.-T. (2012). Effects of aging and education on false memory. *International Journal of Aging & Human Development*, *74*, 287–98. doi:10.2190/AG.74.4.b

*Lee, H. N., Rosa, N. M., & Gutchess, A. H. (2016). Ageing and the group-reference effect in memory. *Memory*, *24*(6), 746–756. doi:10.1080/09658211.2015.1049184

*Light, L. L., & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 531–541. doi:10.3389/fpsyg.2013.00639

Lineweaver, T. T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and

aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition*, *5*, 264-296.

Lockhart, R. S., & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, *74*, 100-109.

Luo, L., & Craik, F. I. (2008). Aging and memory: A cognitive approach. *The Canadian Journal of Psychiatry*, *53*, 346-353.

*Luo, L., Hendriks, T., & Craik, F. I. M. (2007). Age differences in recollection: Three patterns of enhanced encoding. *Psychology and Aging*, *22*, 269–280. doi:10.1037/0882-7974.22.2.269

*Ly, M., Murray, E., & Yassa, M. A. (2013). Perceptual versus conceptual interference and pattern separation of verbal stimuli in young and older adults. *Hippocampus*, *23*(6), 425–430. doi:10.1002/hipo.22110

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory* (2nd ed.). New York: Erlbaum.

*Macpherson, H., Pipingas, A., & Silberstein, R. (2009). A steady state visually evoked potential investigation of memory and ageing. *Brain and Cognition*, *69*, 571–579. doi:10.1016/j.bandc.2008.12.003

*Mason, S. E. (1986). Age and gender as factors in facial recognition and identification. *Experimental Aging Research*, *12*, 151–4. doi:10.1080/03610738608259453

Mather, M. (2004). Aging and emotional memory. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 272-307). New York: Oxford University Press.

*Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, *16*, 437–458.

*Matzen, L. E., & Benjamin, A. S. (2013). Older and wiser: Older adults' episodic word memory benefits from sentence study contexts. *Psychology and Aging*, *28*, 754 -767.

*May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age

differences in memory. Psychological Science, 4(5), 326-330. doi: 10.1111/j.1467-
9280.1993.tb00573.x

*May, C. P., Rahhal, T., Berry, E. M., & Leighton, E. A. (2005). Aging, source memory, and emotion.
Psychology and Aging, 20, 571–578. doi:10.1037/0882-7974.20.4.571

McGillivray, S., & Castel, A. D. (2017). Older and younger adults' strategic control of metacogntiive
monitoring: The role of consequences, task experience, and prior knowledge. *Experimental Aging
Research*, *43*, 233-256. doi: 10.1080/0361073X.2017.1298956

*McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in
item recognition. *Journal of Memory and Language*, *66*, 416–437. doi:10.1016/j.jml.2011.12.001

*McKoon, G., & Ratcliff, R. (2013). Aging and predicting inferences: A diffusion model analysis.
Journal of Memory and Language, *68*, 240–254. doi:10.1016/j.jml.2012.11.002

McShane, B. B., & Böckenholt, U. (2014). You cannot step into the same river twice: When power
analyses are optimistic. *Perspectives on Psychological Science*, *9*, 612-625. doi:
10.1177/1745691614548513

McShane, B.B., Böckenholt, U., & Hansen, K.T. (2016). Adjusting for publication bias in meta-
analysis: An evaluation of selection methods and some caution. *Perspectives on Psycholgoical
Science*, *11*, 730-749. doi:10.1177/1745691616662243

*Meade, M. L., & Roediger, H. L. (2009). Age differences in collaborative memory: the role of
retrieval manipulations. *Memory & Cognition*, *37*, 962–975. doi:10.3758/MC.37.7.962

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman:
Benefits for lifelong learning. *Psychology and Aging*, *28*, 142-147. doi: 10.1037/a0030890

*Mickley, K. R., & Kensinger, E. A. (2009). Phenomenological characteristics of emotional memories
in younger and older adults. *Memory*, *17*, 528–543. doi:10.1080/09658210902939363

*Mohanty, P., Naveh-Benjamin, M., & Ratneshwar, S. (2016). Beneficial Effects of Semantic Memory

Support on Older Adults' Episodic Memory: Differential Patterns of Support of Item and Associative Information, *31*(1), 25–36. doi:10.1037/pag0000059

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The PRISMA Group. (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. *PLOS Medicine*, *6*, e1000097. doi:10.1371/journal.pmed1000097

*Morcom, A. M., Bullmore, E. T., Huppert, F. A., Lennox, B., Praseedom, A., Linnington, H., & Fletcher, P. C. (2010). Memory encoding and dopamine in the aging brain: A psychopharmacological neuroimaging study. *Cerebral Cortex*, *20*, 743–757. doi:10.1093/cercor/bhp139

*Morcom, A. M., Good, C. D., Frackowiak, R. S. J., & Rugg, M. D. (2003). Age effects on the neural correlates of successful memory encoding. *Brain*, *126*, 213–229. doi:10.1093/brain/awg020

*Morcom, A. M., Li, J., & Rugg, M. D. (2007). Age effects on the neural correlates of episodic retrieval: Increased cortical recruitment with matched performance. *Cerebral Cortex*, *17*, 2491–2506. doi:10.1093/cercor/bhl155

*Mormino, E. C., Brandel, M. G., Madison, C. M., Marks, S., Baker, S. L., & Jagust, W. J. (2012). Aβ Deposition in aging is associated with increases in brain activation during successful memory encoding. *Cerebral Cortex*, *22*(8), 1813–1823. doi:10.1093/cercor/bhr255

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*, 1170-1187. doi:10.1037/0278-7393.29.5.1170

*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*, *29*, 826–837. doi:10.1037/0278-7393.29.5.826

*Naveh-Benjamin, M., & Kilb, A. (2012). How the measurement of memory processes can affect

memory performance: The case of remember/know judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 194–203. doi:10.1037/a0025256

*Naveh-Benjamin, M., & Kilb, A. (2014). Age-related differences in associative memory: The role of sensory decline. *Psychology and Aging*, *29*(3), 672–683. doi:10.1037/a0037138

*Naveh-Benjamin, M., Maddox, G. B., Jones, P., Old, S., & Kilb, A. (2012). The effects of emotional arousal and gender on the associative memory deficit of older adults. *Memory and Cognition*, *40*, 551–566. doi:10.3758/s13421-011-0169-x

*Naveh-Benjamin, M., Shing, Y. L., Kilb, A., Werkle-Bergner, M., Lindenberger, U., & Li, S.-C. (2009). Adult age differences in memory for name-face associations: the effects of intentional and incidental learning. *Memory*, *17*, 220–232. doi:10.1080/09658210802222183

Naveh-Benjamin, M., & Smyth, A. C. (2016). DRYAD and ADH: Further comments on explaining age-related differences in memory. *Psychology and Aging*, *31*, 21-24. doi: 10.1037/pag0000066

*Obermeyer, S., Kolling, T., Schaich, A., & Knopf, M. (2012). Differences between old and young adults' ability to recognize human faces underlie processing of horizontal information. *Frontiers in Aging Neuroscience*, *4*, 1–9. doi:10.3389/fnagi.2012.00003

Old, S. R., & Naveh-Benjamin, M. (2008). Memory for people and their actions: further evidence for an age-related associative deficit. *Psychology and Aging*, *23*, 467.

*Olfman, D., Light, L. L., Schmalstig, M., Pospisil, D. A., Pendergrass, R., & Chung, C. (2017). Preserved criterion flexibility in itemrRecognition in older adults. *Psychology and Aging*, *32*(7), 675–680.

Open Science Framework. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, a4716. doi: 10.1126/science.aac4716

*Pachur, T., Mata, R., & Schooler, L. J. (2009). Cognitive aging and the adaptive use of recognition in decision making. *Psychology and Aging*, 24, 901–915. doi:10.1037/a0017211

*Park, D. C., & Puglisi, J. T. (1985). Older adults' memory for color of pictures and words. *The Journal of Gerontology*, *40*, 198–204. doi: 10.1093/geronj/40.2.198

*Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: evidence for pictorial superiority. *The Journal of Gerontology*, *38*, 582–588. doi: 10.1093/geronj/38.5.582

*Parks, C. M., DeCarli, C., Jacoby, L. L., & Yonelinas, A. P. (2010). Aging effects on recollection and familiarity: the role of white matter hyperintensities. *Aging, Neuropsychology, and Cognition*, *17*, 422–438. doi:10.1080/13825580903469838

Parks, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, *73*, 44-58.

*Perfect, T. J., & Moon, H. C. (2005). The own-age effect in face recognition. In J. Duncan, L. Phillips, & P. McLeod (Eds.), *Measuring the mind: Speed, control, and age* (pp. 317–337). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198566427.003.0013

*Peters, J., & Daum, I. (2008). Differential effects of normal aging on recollection of concrete and abstract words. *Neuropsychology*, *22*, 255–261. doi:10.1037/0894-4105.22.2.255

Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association*, *296*, 676-680.

Peters, J.L., Sutton, A.J, Jones, D.R., Abrams, K.R., Rushton, L., & Moreno, S.G. (2010). Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. *Journal of the Royal Statistical Society*, *Series A*, 173, 575-591.

*Pezdek, K. (1987). Memory for pictures: a life-span study of the role of visual detail. *Child Development*, *58*, 807–815. doi:10.2307/1130218

*Pierce, B. H., Waring, J. D., Schacter, D. L., & Budson, A. E. (2008). Effects of distinctive encoding

on source-based false recognition: further examination of recall-to-reject processes in aging and Alzheimer disease. *Cognitive and Behavioral Neurology*, *21*, 179–186. doi:10.1097/WNN.0b013e31817d74e7

*Pitarque, A., Meléndez, J. C., Sales, A., Mayordomo, T., Satorres, E., Escudero, J., & Algarabel, S. (2016). The effects of healthy aging, amnestic mild cognitive impairment, and Alzheimer's disease on recollection, familiarity and false recognition, estimated by an associative process-dissociation recognition procedure. *Neuropsychologia*, *91*, 29–35. doi:10.1016/j.neuropsychologia.2016.07.010

Pratte, M. S., & Rouder, J. N. (2012). Assessing the dissociability of recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1591-1607. doi: 10.1037/a0028144

*Prull, M. W., Dawes, L. L. C., Martin, A. M., Rosenberg, H. F., & Light, L. L. (2006). Recollection and familiarity in recognition memory: adult age differences and neuropsychological test correlates. *Psychology and Aging*, *21*, 107–118. doi:10.1037/0882-7974.21.1.107

*Prull, M. W., Light, L. L., Collett, M. E., & Kennison, R. F. (1998). Age-related differences in memory illusions: revelation effect. *Aging, Neuropsychology, and Cognition*, *5*, 147–165. doi:10.1076/anec.5.2.147.598

*Rabinowitz, J. C. (1989). Judgments of origin and generation effects: comparisons between young and elderly adults. *Psychology and Aging*, *4*, 259–268. doi:10.1037/0882-7974.4.3.259

Rabinowitz, J. C., & Craik, F. I. M. (1986). Prior retrieval effects in young and old adults. *Journal of Gerontology*, *41*, 368-375. doi.org/10.1093/geronj/41.3.368

Rahhal, T. A., Hasher, L., & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: now you see them, now you don't. *Psychology and Aging*, *16*, 697.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline:

Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*, 5-42.

*Rankin, J. L., & Kausler, D. H. (1979). Adult age differences in false recognitions. *The Journal of Gerontology*, *34*, 58–65. doi: 10.1093/geronj/34.1.58

*Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*, 408–424. doi:10.1016/j.jml.2003.11.002

*Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*, 464-487. doi: 10.1037/a0023810

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin*, *138*, 146-174. doi: 10.1037/a0025750

*Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging*, *23*, 239–249. doi:10.1037/0882-7974.23.2.239

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

*Rosa, N. M., & Gutchess, A. H. (2013). False memory in aging resulting from self-referential processing. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *68*(6), 882–892. doi:10.1093/geronb/gbt018

Rotello, C. M., & Macmillan, N. A. (2006). Remember–know models as decision strategies in two experimental paradigms. *Journal of Memory and Language*, *55*, 479-494.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734-760.

*Rybarczyk, B. D., Hart, R. P., & Harkins, S. W. (1987). Age and forgetting rate with pictorial stimuli. *Psychology and Aging*, *2*, 404–406. doi:10.1037/0882-7974.2.4.404

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition.

    Psychological Review, *103*, 403-428.

*Saverino, C., Fatima, Z., Sarraf, S., Oder, A., Strother, S. C., & Grady, C. L. (2016). The associative

    memory deficit in aging is related to reduced selectivity of brian activity during encoding. *Journal*

    of Cognitive Neuroscience, *28*(9), 1331–1344. doi:10.1162/jocn_a_00970

*Schacter, D. L., Cooper, L. A., & Valdiserri, M. (1992). Implicit and explicit memory for novel visual

    objects in older and younger adults. *Psychology and Aging*, *7*, 299–308. doi:10.1037/0882-

    7974.7.2.299

*Schefter, M., Knorr, S., Kathmann, N., & Werheid, K. (2012). Age differences on ERP old/new effects

    for emotional and neutral faces. *International Journal of Psychophysiology*, *85*, 257–269.

    doi:10.1016/j.ijpsycho.2011.11.011

*Scheuplein, A. L., Bridger, E. K., & Mecklinger, A. (2014). Is faster better? Effects of response

    deadline on ERP correlates of recognition memory in younger and older adults. *Brain Research*,

    1582, 139–153. doi:10.1016/j.brainres.2014.07.025

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of*

    Psychology/Revue canadienne de psychologie, *20*, 228-236.

*Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in

    eyewitness identification and face recognition. *Memory & Cognition*, *27*, 538–552.

    doi:10.3758/BF03211547

*Simons, J. S., Dodson, C. S., Bell, D., & Schacter, D. L. (2004). Specific- and partial-source memory:

    effects of aging. *Psychology and Aging*, *19*, 689–694. doi:10.1037/0882-7974.19.4.689

*Skinner, E. I., & Fernandes, M. A. (2008). Interfering with remembering and knowing: Effects of

    divided attention at retrieval. *Acta Psychologica*, 1*27*, 211–221. doi:10.1016/j.actpsy.2007.05.001

*Skinner, E. I., & Fernandes, M. A. (2009). Illusory recollection in older adults and younger adults

under divided attention. *Psychology and Aging*, *24*, 211–216. doi:10.1037/a0014177

Smyth, A. C., & Naveh-Benjamin, M. (2016). Can DRYAD explain age-related associative memory

deficits?. *Psychology and Aging*, *31*, 1-13. doi: 10.1037/a0039071

*Soei, E., & Daum, I. (2008). Course of relational and non-relational recognition memory across the

adult lifespan. *Learning & Memory*, *15*, 21–28. doi:10.1101/lm.757508

*Souchay, C., Moulin, C. J. A., 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*, 769–784. doi:10.1016/j.concog.2006.11.002

*Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in

episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 32, 101–117. doi:10.1037/0278-7393.32.1.101

*Spaniol, J., Voss, A., & Grady, C. L. (2008). Aging and emotional memory: cognitive mechanisms

underlying the positivity effect. *Psychology and Aging*, *23*, 859–872. doi:10.1037/a0014218

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: a

meta-analysis. *Psychology and Aging*, *10*, 527-539.

*Springer, M. V., McIntosh, A. R., Winocur, G., & Grady, C. L. (2005). The relation between brain

activity during memory tasks and years of education in young and older adults. *Neuropsychology*,

19, 181–192. doi:10.1037/0894-4105.19.2.181

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African

Americans. *Journal of Personality and Social Psychology*, *69*, 797-811. doi: 10.1037/0022-

3514.69.5.797

Stram, D.O. (1996). Meta-analysis of published data using a linear mixed-effects model. *Biometrics*,

52, 536-544.

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological*

Review, *19*, 447-462. doi: 10.1037/h0069812

*Suengas, A. G., Ruiz Gallego-Largo, T., & Simón, T. (2010). Age-related changes in recognition and response criterion. *The Spanish Journal of Psychology*, *13*, 557–571.

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *The Quarterly Journal of Experimental Psychology*, *69*, 1876-1889. doi: 10.1080/17470218.2016.1146780

*Tauber, S. K., Dunlosky, J., Urry, H. L., & Opitz, P. C. (2017). The effects of emotion on younger and older adults' monitoring of learning. *Aging, Neuropsychology, and Cognition*, *24*(5), 555–574. doi:10.1080/13825585.2016.1227423

*Téllez-Alanís, B., & Cansino, S. (2004). Incidental and intentional encoding in young and elderly adults. *NeuroReport*, *15*, 1819–1823. doi:10.1097/01.wnr.0000137075.41257.98

*Thapar, A., & Sniezek, S. M. (2008). Aging and the revelation effect. *Psychology and Aging*, *23*, 473–477. doi:10.1037/0882-7974.23.2.473

*Thapar, A., & Westerman, D. L. (2009). Aging and fluency based illusions in recognition memory. *Psychology and Aging*, *24*, 595–603. doi:10.1037/a0016575

*Thomas, A. K., & Dubois, S. J. (2011). Reducing the burden of stereotype threat eliminates age differences in memory distortion. *Psychological Science*, *22*, 1515–1517. doi:10.1177/0956797611425932

*Thomas, R. C., & Hasher, L. (2006). The influence of emotional valence on age differences in early processing and memory. *Psychology and Aging*, *21*, 821–5. doi:10.1037/0882-7974.21.4.821

*Thomas, A. K., & McDaniel, M. A. (2013). The interaction between frontal functioning and encoding processes in reducing false memories. *Aging, Neuropsychology, and Cognition*, *20*(4), 443–470. doi:10.1080/13825585.2012.736468

*Till, R. E., Bartlett, J. C., & Doyle, A. H. (1982). Age differences in picture memory with resemblance

and discrimination tasks. *Experimental Aging Research*, *8*, 179–184.

doi:10.1080/03610738208260362

*Toth, J. P., Daniels, K. A., & Solinger, L. A. (2011). What you know can hurt you: Effects of age and

prior knowledge on the accuracy of judgments of learning. *Psychology and Aging*, *26*, 919–31.

doi:10.1037/a0023379

*Trelle, A. N., Henson, R. N., Green, D. A. E., & Simons, J. S. (2017). Declines in representational

quality and strategic retrieval processes contribute to age-related increases in false recognition.

Journal of Experimental Psychology: Learning Memory and Cognition, *43*(12), 1883–1897.

doi:10.1037/xlm0000412

*Trott, C., T. Friedman, D., Ritter, W., Fabiana, M., & Snodgrass, J. G. (1999). Episodic priming and

memory for temporal source: event-related potentials reveal age-related differences in

prefrontal functioning. *Psychology and Aging*, *14*, 390–413.

*Troyer, A. K., Hafliger, A., Cadieux, M. J., & Craik, F. I. M. (2006). Name and face learning in older

adults: effects of level of processing, self-generation, and intention to learn. *The Journals of

Gerontology Series B: Psychological Sciences and Social Sciences*, *61B*, 67–74.

doi:10.1093/geronb/61.2.P67

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory

and Language*, *64*, 109-118.

*Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the

elderly: evidence from metamnemonic judgments of word frequency. *Psychology and Aging*, *27*,

683–90. doi:10.1037/a0025838

*Vakil, E., Hornik, C., & Levy, D. A. (2008). Conceptual and perceptual similarity between encoding

and retrieval contexts and recognition memory context effects in older and younger adults. *The

Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63B*, 171–175.

*Vakil, E., Raz, T., & Levy, D. A. (2010). Probing the brain substrates of cognitive processes

    responsible for context effects on recognition memory. *Aging, Neuropsychology, and Cognition*,

    17, 519–544. doi:10.1080/13825581003690182

*Velanova, K., Lustig, C., Jacoby, L. L., & Buckner, R. L. (2007). Evidence for frontally mediated

    controlled processing differences in older adults. *Cerebral Cortex*, *17*, 1033–1046.

    doi:10.1093/cercor/bhl013

Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A

    quantitative integration of research findings. *The Journal of Gerontology*, *48*, P157-P171. doi:

    10.1093/geronj/48.4.P157

*Wang, T. H., de Chastelaine, M., Minton, B., & Rugg, M. D. (2012). Effects of age on the neural

    correlates of familiarity as indexed by ERPs. *Journal of Cognitive Neuroscience*, 24(5), 1055-

    1068. doi: 10.1162/jocn_a_00129

*Wang, T. H., Johnson, J. D., de Chastelaine, M., Donley, B. E., & Rugg, M. D. (2016). The effects of

    age on the neural correlates of recollection success, recollection-related cortical reinstatement, and

    post-retrieval ,onitoring. *Cerebral Cortex*, *26*(4), 1698–1714. doi:10.1093/cercor/bhu333

*Waring, J. D., & Kensinger, E. A. (2009). Effects of emotional valence and arousal upon memory

    trade-offs with aging. *Psychology and Aging*, *24*, 412–422. doi:10.1037/a0015526

*Werheid, K., Gruno, M., Kathmann, N., Fischer, H., Almkvist, O., & Winblad, B. (2010). Biased

    recognition of positive faces in aging and amnestic mild cognitive impairment. *Psychology and*

    Aging, *25*, 1–15. doi:10.1037/a0018358

*Wiese, H., Komes, J., & Schweinberger, S. R. (2012) Daily-life contact affects the own-age bias and

    neural correlates of face memory in elderly participants. *Neuropsychologia*, *50*, 3496–3508.

    doi:10.1016/j.neuropsychologia.2012.09.022

*Wiese, H., Schweinberger, S. R., & Hansen, K. (2008). The age of the beholder: ERP evidence of an

own-age bias in face memory. *Neuropsychologia*, *46*, 2973–2985.

doi:10.1016/j.neuropsychologia.2008.06.007

*Wiggs, C. L., & Martin, A. (1994). Aging and feature-specific priming of familiar and novel stimuli.

Psychology and Aging, *9*, 578–88. doi:10.1037/0882-7974.9.4.578

Wixted, J. T. (1998). Remembering and forgetting. In K. A. Lattal & M. Perone (Eds.), *Handbook of*

research methods in human operant behavior (pp. 263-289). New York: Springer.

Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of

retrograde amnesia. *Psychological Review*, *111*, 864-879. doi: 10.1037/0033-295X.111.4.864

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory.

Psychological Review, *114*, 152-176. doi: 10.1037/0033-295X.114.1.152

*Wolff, N., Wiese, H., & Schweinberger, S. R. (2012). Face recognition memory across the adult life

span: Event-related potential evidence from the own-age bias. *Psychology and Aging*, *27*, 1066–

1081. doi:10.1037/a0029112

*Wolk, D. A., Sen, N. M., Chong, H., Riis, J. L., McGinnis, S. M., Holcomb, P. J., & Daffner, K. R.

(2009). ERP correlates of item recognition memory: Effects of age and performance. *Brain*

Research, *1250*, 218–231. doi:10.1016/j.brainres.2008.11.014

*Yang, L., Truong, L., Fuss, S., & Bislimovic, S. (2012). The effects of ageing and divided attention on

the self-reference effect in emotional memory: Spontaneous or effortful mnemonic benefits?

Memory, *20*, 596–607. doi:10.1080/09658211.2012.690040

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-

process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*,

1341-1354. doi: 10.1037/0278-7393.20.6.1341

Zacks, R. T., & Hasher, L. (2006). Aging and long-term memory: Deficits are not inevitable. In E.

Bialystok and F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 162–177).

New York: Oxford University Press.

Appendix A

As a supplement to our primary analysis, we also modeled age differences in hit rate (which reflects a combination of greater discrimination ability and a liberal criterion) and false alarm rate (which reflects a combination of poorer discrimination ability and a liberal criterion). Specifically, we used the differences in the log odds (logit; Agresti, 2007) of hit responses and of false alarms responses to avoid problems with the addition or subtraction of proportions. Aside from the change in the dependent variable, the models were otherwise identical to those reported in the primary analysis.

These results, reported in Table A1 and Table A2, largely confirm the results of the main analysis. For instance, older adults had both an overall lower hit rate than young adults, $t = -8.94$, $p < .001$, and a higher false alarm rate, $t = 15.20$, $p < .001$, consistent with the finding that older adults had poorer overall discrimination. Further, age differences were larger in false alarms: The odds of false alarms were 1.65 times greater for older adults than young adults, but the odds of hits were only 1.39 times greater for young adults than older adults. This accords with the finding that older adults overall had a somewhat more liberal response criterion (which would increase the false alarm rate) than young adults.

Some effects on discrimination were evident in both hit rates and false alarms (i.e., a mirror effect; Glanzer, Adams, Iverson, & Kim, 1993): Longer retention intervals both lowered older adults' hit rates and increased their false alarms relative to young adults. Some differences in criterion placement were also observed across hit and false alarms. Older adults had both higher hit rates and higher false alarm rates for faces and pictures as compared to words; however, these two effects were also accompanied by elevated false alarm rates, consistent with the liberal criterion shift reported in the main text.

However, some effects were observed only in hit rates or only in false alarms. One particularly striking pattern, as also implied by the changes in *d'* and *c,* is that older adults had an especially

elevated false alarm rate when lures bore some semantic relation to the targets or when the lures contained one component of a studied target, but they had no corresponding change in hit rate. False alarms were also especially elevated for conjunction lures comprised of two studied pieces re-paired although the presence of these lures also led to some decrease in older adults' hit rates. Conversely, older adults' relatively liberal criteria given longer study lists appears to be reflected mostly in hits, and older adults' relatively spared memory for texts was driven largely by hits to previously-presented texts.

As suggested by the absence of any corresponding effects on d' and c, neither the age difference in hit rates nor in false alarm rates was influenced by stimulus valence, presentation modality, production or self-generation at study, shallow encoding tasks, or multiple study opportunities.

Table A1.
Fixed Effect Estimates for Multi-Level Model of Differences Between Age Groups in Log Odds of Hits.

Fixed effect	$\hat{\beta}$	SE	95% CI	χ^2	p
Intercept (baseline age difference)	-0.329	0.037	[-0.401, -0.257]		
Younger adult logit hit rate (slope)	-0.286	0.028	[-0.341, -0.232]	102.56	< .001
Subject characteristics					
Mean age of older adult group	0.001	0.009	[-0.016, 0.019]	0.02	.89
Stimulus characteristics					
Stimuli are faces (vs. words)	0.283	0.078	[0.129, 0.436]	13.43	< .001
Stimuli are pictures (vs. words)	0.246	0.054	[0.139, 0.352]	19.96	< .001
Stimuli are texts (vs. words)	0.295	0.084	[0.131, 0.459]	12.67	< .001
Positive valence	0.028	0.066	[-0.101, 0.156]	0.19	.67
Negative valence	-0.039	0.060	[-0.156, 0.079]	0.44	.51
Study phase characteristics					
Auditory presentation (vs. visual)	-0.117	0.097	[-0.308, 0.073]	1.51	.22
Bimodal presentation	0.047	0.200	[-0.345, 0.439]	0.06	.81
Production	-0.045	0.199	[-0.434, 0.344]	0.07	.80
Intentional encoding	-0.004	0.070	[-0.141, 0.134]	< 0.01	.97
Self-paced study	-0.139	0.109	[-0.353, 0.075]	1.75	.19
Self-paced study x intentional encoding	0.286	0.165	[-0.038, 0.610]	3.14	.08
Study time per presentation (s)	-0.023	0.011	[-0.045, -0.000]	4.06	.04
Deep orienting task	-0.114	0.068	[-0.247, 0.020]	2.98	.08
Shallow orienting task	-0.023	0.084	[-0.188, 0.142]	0.12	.73
Generation task	0.113	0.120	[-0.122, 0.349]	0.94	.33
Divided attention	-0.342	0.233	[-0.799, 0.115]	2.28	.13
Number of memoranda	0.001	0.001	[0.001, 0.002]	7.19	.01
Multiple study opportunities	0.103	0.075	[-0.044, 0.250]	1.81	.18
Test phase characteristics					
Intervening cued recall	-0.065	0.186	[-0.430, 0.300]	0.13	.72
Intervening free recall	0.114	0.105	[-0.091, 0.320]	1.31	.25
Continuous recognition	-0.211	0.187	[-0.577, 0.155]	1.43	.23
Retention interval	-0.031	0.009	[-0.048, -0.014]	13.16	< .001
Continuous recognition x retention interval	-0.066	0.040	[-0.144, 0.013]	2.66	.10
Number of test probes	0.001	< 0.001	[0.001, 0.002]	1.47	.23
Conjunction lures	-0.112	0.056	[-0.222, -0.003]	4.21	.04
Component lures	0.102	0.086	[-0.067, 0.272]	1.48	.22
Semantically related lures	0.027	0.046	[-0.063, 0.117]	0.37	.54
Featurally related lures	-0.066	0.059	[-0.181, 0.049]	1.34	.25
Proportion of lures	-0.355	0.219	[-0.784, 0.074]	2.70	.10

Note. SE = standard error. Positive parameter estimates indicate age differences smaller than the mean (i.e., older adults relatively less disadvantaged in hit rates); negative parameter estimates indicate larger age differences than the mean (i.e., older adults were especially disadvantaged).

Table A2.
Fixed Effect Estimates for Multi-Level Model of Differences Between Age Groups in Log Odds of False Alarms.

Fixed effect	$\hat{\beta}$	SE	95% CI	χ^2	p
Intercept (baseline age difference)	0.503	0.033	[0.438, 0.567]		
Younger adult logit false alarm rate (slope)	-0.251	0.025	[-0.300, -0.202]	96.42	< .001
Subject characteristics					
Mean age of older adult group	0.014	0.009	[-0.004, 0.032]	2.36	.12
Stimulus characteristics					
Stimuli are faces (vs. words)	0.522	0.091	[0.344, 0.700]	34.09	.001
Stimuli are pictures (vs. words)	0.152	0.064	[0.028, 0.277]	5.66	.02
Stimuli are texts (vs. words)	-0.083	0.096	[-0.271, 0.105]	0.77	.38
Positive valence	0.120	0.078	[-0.033, 0.274]	2.38	.12
Negative valence	0.072	0.070	[-0.066, 0.210]	1.03	.31
Study phase characteristics					
Auditory presentation (vs. visual)	0.049	0.110	[-0.166, 0.264]	0.19	.66
Bimodal presentation	0.326	0.219	[-0.103, 0.755]	2.37	.12
Production	-0.251	0.212	[-0.667, 0.165]	1.77	.18
Intentional encoding	0.073	0.078	[-0.080, 0.227]	0.92	.34
Self-paced study	0.151	0.124	[-0.092, 0.394]	1.55	.21
Self-paced study x intentional encoding	0.227	0.191	[-0.148, 0.602]	1.59	.21
Study time per presentation (s)	0.002	0.012	[-0.022, 0.026]	0.04	. 85
Deep orienting task	0.054	0.074	[-0.090, 0.198]	0.56	.45
Shallow orienting task	0.093	0.092	[-0.087, 0.273]	0.90	.34
Generation task	-0.078	0.141	[-0.355, 0.199]	0.33	.57
Divided attention	-0.499	0.277	[-1.043, 0.044]	3.46	.06
Number of memoranda	< 0.001	0.001	[-0.001, 0.002]	0.76	.38
Multiple study opportunities	-0.029	0.084	[-0.193, 0.135]	0.15	.69
Test phase characteristics					
Intervening cued recall	0.282	0.208	[-0.126, 0.691]	1.91	.17
Intervening free recall	-0.226	0.121	[-0.463, 0.010]	3.35	.07
Continuous recognition	-0.054	0.182	[-0.411, 0.303]	0.11	.74
Retention interval	0.034	0.010	[0.015, 0.054]	12.29	< .001
Continuous recognition x retention interval	-0.019	0.047	[-0.111, 0.072]	0.16	.69
Number of test probes	> -0.001	0.001	[-0.002, 0.000]	2.36	.12
Conjunction lures	0.408	0.070	[0.271, 0.545]	34.24	< .001
Component lures	0.359	0.103	[0.157, 0.561]	12.53	< .001
Semantically related lures	0.384	0.057	[0.271, 0.496]	44.16	< .001
Featurally related lures	0.021	0.071	[-0.118, 0.161]	0.06	.80
Proportion of lures	-0.411	0.254	[-0.908, 0.086]	2.36	.12

Note. SE = standard error. Positive parameter estimates indicate age differences larger than the mean (i.e., older adults even more apt to false alarm); negative parameter estimates indicate age differences smaller than the mean (i.e., older adults relatively less apt to false alarm).

Appendix B

The regression equation for the model of d' took the following form:

$y_{ijklm} = \gamma_{00000} + \gamma_{10000}(x_{1ijklm} - \bar{x}_1) + \gamma_{20000}x_{2ijklm} + \ldots + \gamma_{31000}x_{31ijklm} + t_{0jklm} + u_{0jklm} + v_{00klm} + w_{000lm} +$

e_{ijklm}

where y_{ijklm} is the d' difference score in condition i between paired older adults and young adult subject

groups j from experiment k in paper l from laboratory m, γ_{00000} is the intercept (equaling the grand

mean d' difference score across all conditions), γ_{10000} is the slope relating younger adults' d' scores to

the difference score, x_{1ijklm} is the corresponding younger adult d' score, $\bar{x}_1$ is the mean younger adult d'

across all conditions in all experiments, γ_{20000} through γ_{310000} are the effects of each of the $n-1$ coded

stimulus and task variables described in the Method section, x_{2ijklm} through x_{nijklm} are the values of each

of those stimulus and task variables in condition i, t_{0jklm} is the random effect of subject group i

(independently sampled from a normal distribution of subject group effects with mean 0 and variance

τ_{T0}^2), u_{00klm} is the random effect of experiment k (independently sampled from a normal distribution of

experiment effects with mean 0 and variance τ_{U00}^2), v_{000lm} is the random effect of paper l (independently

sampled from a normal distributed of paper effects with mean 0 and variance τ_{V000}^2), w_{0000m} is is the

random effect of laboratory m (independently sampled from a normal distributed of laboratory effects

with mean 0 and variance τ_{W0000}^2), and e_{ijklm} is a random error term (independently sampled from a

normal distribution with mean 0 and variance $\sigma_{\hat{e}}^2$). Except where otherwise noted in the Method

section, each continuous x variable was centered around its mean and the categorical x variables were

coded using effects coding so that the intercept term corresponded to the grand mean across all

conditions.

Observations were weighted with weights equal to the combined sample size (older plus young

adults) in condition i.

The equation for the model of c was identical except that the measures of older and young

adults' d' were replaced by measures of older and young adults' c.