Younger and older adults weigh multiple cues in a similar manner to generate judgments of learning

By: Jarrod C. Hines*, Christopher Hertzog† & Dayna R. Touron‡

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

This is an Accepted Manuscript of an article published by Taylor & Francis in *Aging, Neuropsychology, and Cognition* on 01 Apr 2015, available online: https://doi.org/10.1080/13825585.2015.1028884

Abstract:

One's memory for past test performance (MPT) is a key piece of information individuals use when deciding how to restudy material. We used a multi-trial recognition memory task to examine adult age differences in the influence of MPT (measured by actual Trial 1 memory accuracy and subjective confidence judgments, CJs) along with Trial 1 judgments of learning (JOLs), objective and participant-estimated recognition fluencies, and Trial 2 study time on Trial 2 JOLs. We found evidence of simultaneous and independent influences of multiple objective and subjective (i.e., metacognitive) cues on Trial 2 JOLs, and these relationships were highly similar for younger and older adults. Individual differences in Trial 1 recognition accuracy and CJs on Trial 2 JOLs indicate that individuals may vary in the degree to which they rely on each MPT cue when assessing subsequent memory confidence. Aging appears to spare the ability to access multiple cues when making JOLs.

Keywords: judgments of learning | memory for past test | aging | metacognition | multilevel modeling

Article:

* Corresponding author. Email: hinesjc@uwec.edu Present address of Jarrod C. Hines is Psychology Department, University of Wisconsin-Eau Claire, Eau Claire, WI, USA.
† School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA
‡ Department of Psychology, University of North Carolina Greensboro, Greensboro, NC, USA

INTRODUCTION

During the course of learning, people are often able to study materials multiple times (e.g., when a student spends multiple nights studying for an exam). Actively monitoring one's current state of learning can help to optimize further study efforts (e.g., Nelson & Narens, 1990; see Bjork, Dunlosky, & Kornell, 2013 for a recent review of self-regulated learning). One metric for assessing an individual's perceptions of his or her current state of learning is the judgment of learning (JOL), for which people rate confidence in the likelihood of remembering a piece of information (e.g., an item from a word list) during a subsequent memory test (for an introductory overview, see Dunlosky & Metcalfe, 2009). As reviewed in detail below, research on aging and JOLs indicates age-related invariance in the relative accuracy of JOLs, as indicated by within-person correlations of item JOLs with memory outcomes such as item recall success (Hertzog & Dunlosky, 2011). The present study focuses on age differences in the information people use to make JOLs and whether the cues that are accessed differ across the adult lifespan (Hertzog & Shing, 2011). Before reviewing the literature on aging and JOLs, we first consider theories of JOL construction that set the stage for understanding age differences in how people use multiple cues to construct JOLs.

Cue utilization and JOLs

It is widely accepted that metacognitive judgments are explicit or implicit inferences based on accessible information, or cues, about underlying cognitive states without direct access to those internal cognitive states (Koriat, 1997) ; Nelson, 1996a) There are many potential metacognitive cues available to individuals, which differ widely in their validity for predicting later remembering (Dunlosky & Metcalfe, 2009).

Research over the last 25 years has identified a large number of cues that are accessed when making JOLs, including initial encoding fluency (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Besken & Mulligan, 2013; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Undorf & Erdfelder, 2013); retrieval fluency (e.g., Benjamin, Bjork, & Schwartz, 1998; Hines, Touron, & Hertzog, 2009); perceived item difficulty (Undorf & Erdfelder, 2013); learners' degree of prior familiarity with to-be-learned items (e.g., Shanks & Serra, 2014); how JOLs are framed (e.g., whether prediction instructions focus on studying or forgetting; Ariel, Hines, & Hertzog, 2014; Finn, 2008); spontaneous use of effective encoding strategies (Hertzog, Sinclair, & Dunlosky, 2010); and, perhaps most commonly, observable item characteristics such as associative relatedness or word concreteness (e.g., Begg et al., 1989; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Koriat, 1997; Rhodes & Castel, 2008).

An open question at present is how many cues individuals can access on a given item, or across a set of items. Most of the existing literature has focused on either one or a small set of potential cues. A multiple-cue utilization perspective (Hertzog, Hines, & Touron, 2013) argues that people can and do access and use multiple cues when making JOLs, although the likelihood of accessing multiple cues during the judgment process will be influenced by factors such as motivation, ability, fatigue, item difficulty, and the capacity of attention and working memory. Certainly, not all available cues are accessed and used when making JOLs. Learners have been found to ignore, neglect, or discount many cues that are highly diagnostic of later remembering, such as an asymmetric direction of associative relatedness of cue–target word pairs (e.g., Koriat & Bjork, 2006) or variation in the effectiveness of instructed encoding strategies (e.g., Hertzog et al., 2009). In some cases, cue-neglect may be associated with overshadowing of one cue by another (e.g., Hertzog et al., 2002).

An important JOL cue in multi-trial learning tasks (involving more than one study–test trial) is memory-for-past-test performance (MPT; Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007, 2008; Hertzog et al., 2013; Serra & Ariel, 2014; Tauber & Rhodes, 2012). JOLs made after the first test are typically strongly related to MPT, which has been argued to seed an MPT heuristic (“if I remembered it before, I’ll remember it again”; Finn & Metcalfe, 2007). Individuals could improve JOL accuracy (or resolution) across study– test trials by solely relying on an MPT heuristic, perhaps at the cost of ignoring other valid cues while generating an underconfidence-with-practice effect, in which JOL magnitudes increasingly underestimate future performance by ignoring future learning (Finn & Metcalfe, 2008). Other cues aside from MPT apparently influence multi-trial JOLs (Ariel & Dunlosky, 2011; Tauber & Rhodes, 2012), although the specific cues responsible for the effects reported in those studies were not identified.

Hertzog et al. (2013) used multilevel modeling procedures (Snijders & Bosker, 2012) to show explicitly that encoding fluency was a source of enhanced JOL–memory correlations at a second study trial, in addition to other cues. As expected from the MPT literature studying associative cued recall, Trial 2 JOLs were strongly influenced by prior recognition memory accuracy or success (the index of MPT). The additional cues influencing JOLs included: (a) item-level recognition confidence judgments (CJs); (b) actual recognition response latency; (c) subjective recognition test fluency (measured by recognition response times, RTs); and (d) encoding fluency (measured by strategy completion RTs) immediately prior to the JOL.

CJs in associative recognition tests are influenced by retrieval fluency and accuracy (Hines et al., 2009), while diverging from both due to fallible metacognitive monitoring. For example, factors like specific recollections of study or retrieval strategies (and their degree of success) or the general degree of clarity regarding those recollections may influence memory confidence in associative recognition tests (e.g., Cohn & Moscovitch, 2007; Parks, 2007). In recognition memory contexts, the less-than-perfect accuracy of CJs, as measured by CJ–recognition memory success correlations, allows for detecting influences of both variables on JOLs.

The role of actual retrieval fluency (correct recognition RTs) in influencing CJs and subsequent JOLs is not surprising, given that RTs are typically faster as one’s level of underlying item-learning increases (e.g., Cerella, Onyper, & Hoyer, 2006; Logan, 1988; Ratcliff, Thapar, & McKoon, 2011). Monitoring this relationship may lead individuals to adopt a fluency heuristic (“if I remembered it quickly, I remembered it accurately”; e.g., Benjamin et al., 1998), which may provide useful information in addition to the MPT heuristic when constructing JOLs. The concurrent and independent influence of subjective retrieval fluency (as measured by estimated recognition RTs) controlling on actual RTs, might be more unexpected, given a literature that generally focuses on metacognition for memory response accuracy but not for response latency.

As with CJs and recognition accuracy, the relation of prior test-trial subjective retrieval fluency on later JOLs probably arises because temporal monitoring in memory tasks are not always well calibrated, instead showing a substantial level of aggregate underestimation of RTs and trial-level divergence between actual and estimated RT (Craik & Hay, 1999; Hertzog, Touron, & Hines, 2007). The influence of prior-trial CJs and estimated retrieval latencies on JOLs, controlling on recognition accuracy and actual recognition RTs, demonstrates the importance of metacognitive monitoring of past events on subsequent JOLs and study time (e.g., Hines et al., 2009; Robinson, Hertzog, & Dunlosky, 2006).

Age differences in JOL construction

It is widely accepted that metacognitive judgments are explicit or implicit inferences based on accessible information, or cues, about underlying cognitive states without direct access to those internal cognitive states (Koriat, 1997; Nelson, 1996a). There are many potential metacognitive cues available to individuals, which differ widely in their validity for predicting later remembering (Dunlosky & Metcalfe, 2009).

Research over the last 25 years has identified a large number of cues that are accessed when making JOLs, including initial encoding fluency (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Besken & Mulligan, 2013; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Undorf & Erdfelder, 2013); retrieval fluency (e.g., Benjamin, Bjork, & Schwartz, 1998; Hines, Touron, & Hertzog, 2009); perceived item difficulty (Undorf & Erdfelder, 2013); learners' degree of prior familiarity with to-be-learned items (e.g., Shanks & Serra, 2014); how JOLs are framed (e.g., whether prediction instructions focus on studying or forgetting; Ariel, Hines, & Hertzog, 2014; Finn, 2008); spontaneous use of effective encoding strategies (Hertzog, Sinclair, & Dunlosky, 2010); and, perhaps most commonly, observable item characteristics such as associative relatedness or word concreteness (e.g., Begg et al., 1989; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Koriat, 1997; Rhodes & Castel, 2008).

An open question at present is how many cues individuals can access on a given item, or across a set of items. Most of the existing literature has focused on either one or a small set of potential cues. A multiple-cue utilization perspective (Hertzog, Hines, & Touron, 2013) argues that people can and do access and use multiple cues when making JOLs, although the likelihood of accessing multiple cues during the judgment process will be influenced by factors such as motivation, ability, fatigue, item difficulty, and the capacity of attention and working memory. Certainly, not all available cues are accessed and used when making JOLs. Learners have been found to ignore, neglect, or discount many cues that are highly diagnostic of later remembering, such as an asymmetric direction of associative relatedness of cue–target word pairs (e.g., Koriat & Bjork, 2006) or variation in the effectiveness of instructed encoding strategies (e.g., Hertzog et al., 2009). In some cases, cue-neglect may be associated with overshadowing of one cue by another (e.g., Hertzog et al., 2002).

An important JOL cue in multi-trial learning tasks (involving more than one study–test trial) is memory-for-past-test performance (MPT; Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007, 2008; Hertzog et al., 2013; Serra & Ariel, 2014; Tauber & Rhodes, 2012). JOLs made after the first test are typically strongly related to MPT, which has been argued to seed an MPT heuristic ("if I remembered it before, I'll remember it again"; Finn & Metcalfe, 2007). Individuals could improve JOL accuracy (or resolution) across study– test trials by solely relying on an MPT heuristic, perhaps at the cost of ignoring other valid cues while generating an underconfidence-with-practice effect, in which JOL magnitudes increasingly underestimate future performance by ignoring future learning (Finn & Metcalfe, 2008). Other cues aside from MPT apparently influence multi-trial JOLs (Ariel & Dunlosky, 2011; Tauber & Rhodes, 2012), although the specific cues responsible for the effects reported in those studies were not identified

Hertzog et al. (2013) used multilevel modeling procedures (Snijders & Bosker, 2012) to show explicitly that encoding fluency was a source of enhanced JOL–memory correlations at a second study trial, in addition to other cues. As expected from the MPT literature studying associative cued recall, Trial 2 JOLs were strongly influenced by prior recognition memory accuracy or success (the index of MPT). The additional cues influencing JOLs included: (a) item-level recognition confidence judgments (CJs); (b) actual recognition response latency; (c) subjective recognition test fluency (measured by recognition response times, RTs); and (d) encoding fluency (measured by strategy completion RTs) immediately prior to the JOL.

CJs in associative recognition tests are influenced by retrieval fluency and accuracy (Hines et al., 2009), while diverging from both due to fallible metacognitive monitoring. For example, factors like specific recollections of study or retrieval strategies (and their degree of success) or the general degree of clarity regarding those recollections may influence memory confidence in associative recognition tests (e.g., Cohn & Moscovitch, 2007; Parks, 2007). In recognition memory contexts, the less-than-perfect accuracy of CJs, as measured by CJ–recognition memory success correlations, allows for detecting influences of both variables on JOLs.

The role of actual retrieval fluency (correct recognition RTs) in influencing CJs and subsequent JOLs is not surprising, given that RTs are typically faster as one's level of underlying item-learning increases (e.g., Cerella, Onyper, & Hoyer, 2006; Logan, 1988; Ratcliff, Thapar, & McKoon, 2011). Monitoring this relationship may lead individuals to adopt a fluency heuristic ("if I remembered it quickly, I remembered it accurately"; e.g., Benjamin et al., 1998), which may provide useful information in addition to the MPT heuristic when constructing JOLs. The concurrent and independent influence of subjective retrieval fluency (as measured by estimated recognition RTs) controlling on actual RTs, might be more unexpected, given a literature that generally focuses on metacognition for memory response accuracy but not for response latency.

As with CJs and recognition accuracy, the relation of prior test-trial subjective retrieval fluency on later JOLs probably arises because temporal monitoring in memory tasks are not always well calibrated, instead showing a substantial level of aggregate underestimation of RTs and trial-level divergence between actual and estimated RT (Craik & Hay, 1999; Hertzog, Touron, & Hines, 2007). The influence of prior-trial CJs and estimated retrieval latencies on JOLs, controlling on recognition accuracy and actual recognition RTs, demonstrates the importance of metacognitive monitoring of past events on subsequent JOLs and study time (e.g., Hines et al., 2009; Robinson, Hertzog, & Dunlosky, 2006).

Age differences in JOL construction

The question of interest for this study was whether there are age differences in multiplecue utilization when making JOLs. The existing literature on age differences in JOLs indicates age-related invariance in the effects of different cues on JOLs, but only a few aging studies have evaluated concurrent influences of multiple cues. There is ample evidence that younger and older adults' JOLs are influenced similarly by experimentally manipulated metacognitive cues and that there are few, if any, age differences in JOL accuracy (or resolution), as measured by within-person correlations of JOLs with memory outcomes (e.g., Dunlosky & Hertzog, 2000; Hertzog et al., 2002, 2010; Robinson et al., 2006). With respect to multiple cue utilization, Hertzog et al. (2010) found that adults of all ages attended to their degree of success in using mediational strategies and assigned higher JOLs to well-encoded items for both related and unrelated items.

However, such findings may not generalize to memory task performance monitoring, as required for MPT effects. Older adults are often less accurate in estimating aggregate recall outcomes within levels of experimental variables (e.g., Price, Hertzog, & Dunlosky, 2008), and are prone to high-confidence recognition memory errors (e.g., Dodson, Bawa, & Krueger, 2007; Hertzog & Touron, 2011; Shing, Werkle-Bergner, Li, & Lindenberger, 2009; Wong, Cramer, & Gallo, 2012). Although Tauber and Rhodes (2012) and Serra and Ariel (2014) found that younger and older adults used the MPT heuristic similarly and appropriately to estimate subsequent JOLs in a cued recall task, it could be the case that attempting to monitor the large set of available cues in recognition memory tasks would exceed older adults' diminished executive control and/or processing capacity (e.g., Hedden & Yoon, 2006; Salthouse, 2004). Furthermore, some aspects of performance monitoring may be impaired in old age, given evidence for age differences in resolution of CJs and estimated retrieval fluency (e.g., Craik & Hay, 1999; Hertzog et al., 2007; Hines et al., 2009). Hence, there are reasons to doubt older adults' ability to monitor multiple cues successfully, especially when potent MPT effects could overshadow other available cues.

Alternatively, Tauber and Rhodes (2012) noted that MPT likely relies not only on explicit recollections of actual test performance, but also on a set of implicit cues that operate in the absence of a high-quality recollection of the testing experience. Given agerelated declines in episodic, but not implicit, memory performance (e.g., Fleischman, Wilson, Gabrieli, Bienias, & Bennet, 2004; for reviews, see Light, 2000; Zacks, Hasher, & Li, 2000), older adults might manifest similar relations of metacognitive cues with JOLs, despite age differences in executive function and working memory capacity.

Current hypotheses

This study used a two study–test trial design in a yes–no associative recognition task (e.g., Naveh-Benjamin, 2000) to evaluate age differences in multiple cue influences on JOLs. Data on younger adults in this experiment were analyzed previously generating evidence of multiple-cue utilization in younger adults (Hertzog et al., 2013). The older adult data have not previously been reported.

Our hypotheses derived from the multiple cue-utilization perspective. We hypothesized that older adults are capable of monitoring multiple cues when making JOLs on a second study–test trial. We predicted that, given the existing evidence for spared resolution and sensitivity of JOLs in old age, multiple cues would act simultaneously and independently for both younger and older adults when estimating Trial 2 JOLs, in patterns consistent with spared MPT accuracy in the presence of multiple-cue influences on JOLs. Critically, our specific empirical hypotheses were that (a) Trial 1 recognition CJs and estimated RTs would predict Trial 2 JOLs independently of Trial 1 recognition memory accuracy (i.e., MPT), and (b) variations in the fluency of self-paced study time at Trial 2 would also influence Trial 2 JOLs, given known encoding fluency effects that have been seen in single study–test trial tasks (e.g., Robinson et al., 2006). These effects, along with robust MPT effects in older adults, would provide compelling evidence for older adults' ability to access multiple cues when making Trial 2 JOLs. Finally, we also expected consistency between Trial 1 and Trial 2 JOLs in both item-level and person-level effects, indicating consistency in perceived item difficulty and stable individual differences in both use of the JOL rating scale and levels of confidence about future learning.

METHOD

Design

The experiment was a 2 (Age: Young, Old) × 2 (Study–Test Trial: 1, 2) mixed design, with age as a between-subjects independent variable and study–test trial as a withinsubjects repeated variable.

Participants

Fifty-one younger adults between the ages of 18 and 25 years (M = 19.2) participated, as did 36 older adults who were between the ages of 60 and 75 years (M = 68.4). All younger adults were undergraduates at the Georgia Institute of Technology who received course credit for their participation. All older adults were recruited from the nearby community and were given an honorarium for their participation. All participants were pre-screened for basic health issues that could impede participation (e.g., diagnosed Parkinson's disease or uncorrected visual impairments). Two cognitive tests were given in addition to the main experimental task to assess the similarity of basic information processing between our younger and older samples. The Shipley Institute of Living Scale– Revised vocabulary test (Zachary, 1986) was given to assess general crystallized abilities and is particularly relevant due to our experimental task involving encoding words (which benefits from knowing their meaning). The Digit–Symbol subtest of the Wechsler Adult Intelligence Scale–Revised (WAIS-R, Wechsler, 1981) was also administered as a test of general fluid ability. This test involves participants being given a sheet of many rows of randomly ordered numbers (one through nine) under which they are to fill in an associated symbol (a key providing the unique associations is included on the top of the fill-in-the blank test). Following the 90-s fill-in-the blank interval, the key and response sheet are removed and participants are given another sheet with a single row of response boxes labeled with numbers one through nine and asked to fill in the appropriate symbol below each number using only their memory. Performance in both portions of the Digit–Symbol subtest benefits from learning the unique associations between the numbers and symbols quickly, an ability that serves as a proxy for general fluid intelligence. Participant characteristics are presented in Table 1. As would be expected given the vast literature on age differences in crystallized and fluid abilities, older adults outperformed younger adults on the Shipley vocabulary measure, whereas younger adults performed better on both components of the WAIS-R Digit–Symbol test. Extremely low outliers (greater than two standard deviations from the mean) would have been culled from the sample, but that was unnecessary. Outcomes pertaining to the other variables in Table 1 are discussed in the Results section.

Table 1. Means and standard deviations of participant characteristics and performance indices for older adults.

	Younger adults		Older adults			
Measure	Mean	*SD*	Mean	*SD*	*t*	*d*
Age (years)	19.22	1.19	68.36	5.44	−62.56***	
Vocabulary	31.41	3.52	34.14	4.02	−3.36**	0.73
Digit-symbol	70.90	10.66	42.91	9.04	12.58***	2.76
Digit-symbol memory	8.14	1.28	4.50	1.80	10.90***	2.39
Trial 1 study time (s)	7.36	4.66	10.00	6.94	−2.12*	0.46
Trial 1 judgment of learning	40.76	17.66	31.45	21.49	2.21*	0.48
Trial 1 recognition accuracy (%)	82.16	12.74	72.36	22.02	2.62*	0.57
Trial 1 confidence judgment	85.53	10.09	76.13	18.36	3.07**	0.67
Trial 1 recognition RT (s)	1.65	0.30	3.55	1.08	−11.93***	2.59
Trial 1 recognition RT estimate (s)	1.49	0.86	1.69	1.38	−0.82	0.18
Trial 2 study time (s)	1.62	2.09	3.50	1.89	−4.31***	0.93
Trial 2 judgment of learning	81.70	13.41	69.33	25.02	2.98**	0.65

Notes: Vocabulary means number of correct out of 40 on the Shipley Institute of Living Scale Revised vocabulary test (Zachary, 1986). Digit-symbol means Wechsler Adult Intelligence Scale-Revised Digit-Symbol Subtest (Wechsler, 1981). Digit-symbol memory means symbol recall memory, following the Wechsler Adult Intelligence Scale – Revised Digit–Symbol subtest. Cohen (1988) advised that values of *d* of 0.2, 0.5, and 0.8 should be considered small, moderate, and large effects, respectively.
*$P < .05$; **$P < .01$; ***$P < .001$.

Materials and procedure

Stimuli and procedural details were identical to those reported by Hertzog et al. (2013). One hundred twenty concrete nouns were used to construct 60 semantically unrelated word pairs (e.g., IVY-BIRD). Participants completed two study–test trials using the same set of 60 word pairs in both trials. Participants were told prior to beginning the first study trial that their memories of each word pair would be tested. Each study–test trial involved randomized item-level study for each word pair that was terminated by participants pressing the spacebar. After ending study for each item, participants gave a JOL relating their ability to remember the word pair approximately 10 min later on a continuous scale from 0% to 100% confidence.

Testing during both trials was also self-paced. Participants were presented with all 60 previously studied (i.e., intact) items in addition to 60 new (i.e., rearranged) items that were constructed from the same 120 nouns used to create the original items (sampled without replacement). For example, IVY-BARREL might have been a new item constructed for testing by rearranging words from the studied word pairs IVY-BIRD and BARREL-STAR. This procedure resulted in 120 test items for which participants were asked to indicate by keypress if they had studied the presented word pair (YES) or if the word pair presented had not been studied previously (NO). After a Yes/No response was registered, participants gave a CJ for their recognition response scaled from 0% to 100%. Participants also gave an RT estimate for their Yes/No keypress using a computer mouse to move a slider along a continuous scale from 0 to 10 s. As the mouse moved, the computer displayed a numeric value to participants that was accurate to the nearest 0.1 s. The order of collecting CJs and RT estimates was counterbalanced across participants who were randomly assigned to counterbalancing conditions to investigate possible reactivity due to the order of metacognitive judgments. No important interactions were found and data were analyzed after collapsing across counterbalancing conditions. Finally, it is unclear how one would map a Trial 1 rearranged-pair test item onto Trial 2 study behavior, given that rearranged test pairs draw words from two different paired-associate items. Moreover, consistent with past research (e.g., Cohn & Moscovitch, 2007; Hertzog & Touron, 2011), relationships between our variables of interest were considerably weaker for rearranged items at test. Hence the inclusion of data from Trial 1 test was restricted to intact word pairs in all analyses.

Statistical procedure

The relationships existing between the outcome variable Trial 2 JOL and predictor variables (specified below) were examined via multilevel modeling with within- and between-person centering of predictors in our models to account for both types of effects (see Enders & Tofighi, 2007; and Singer, 1998). Within-person centering rescales each predictor independently as its deviation from that person's mean value. For example, if a participant's average RT is 2000 ms, an item-level RT of 1500 ms would be scaled as −500 ms. This type of centering allows for the evaluation of the influences of within person variation in cues predicts within-person variation in JOLs. In multilevel regression, these influences are called Level 1 regression effects. Between-person centering involves calculating an average value for each predictor for each participant and calculating a difference score between the mean of each participant and the mean of the overall sample. For example, if the sample mean RT was 3000 ms, and a participant's mean RT was 2000 ms, the value would be rescaled to −1000 ms for every item-level response. Between-person centering allows for the evaluation of individual differences in means on predictors and are referred to as Level 2 regression effects in multilevel models.

Level 1 and Level 2 effects act independently from one another and can be used as predictors within the same model (Singer, 1998). This is advantageous because questions pertaining to within- and between-person variations in predictors can be investigated simultaneously. For example, questions like "does a faster recognition response for an item in Trial 1 predict a higher JOL for that item in Trial 2?" can be asked along with "do people who respond faster on average in Trial 1 have higher average memory confidence in Trial 2?" Examining only one type of question in isolation can lead to improper conclusions about the relationships between predictors and outcomes because Level 1 and Level 2 effects can operate in opposing directions (e.g., people may be more confident in items that they remember faster, but people who, on average, respond quickly are not necessarily more confident than those who respond more slowly). See Snijders and Bosker (2012) for a more detailed treatment of this topic.

The multilevel regression models included the within- and between-person variables associated with every factor measured explicitly during the course of our experiment: participant age group (younger or older), study time from Trials 1 and 2, Trial 1 JOL, recognition accuracy (a binary measure), CJ, RT, and estimated RT. JOLs and CJs were scaled from 0% to 100% memory confidence, time-based measures were scaled in milliseconds, and age interaction terms were included for each variable to afford an age-based comparison of relationships between variables.

Models were fit to our data using SAS PROC MIXED (Littell, Milliken, Stroup, & Wolfinger, 2000), and our Type I error was set at $\alpha = .05$ for all main effects and interactions. The *F*-tests associated with all fixed effects are reported, but only effects that met this standard are discussed. Relevant to our error criterion, it is appropriate to mention another benefit of examining our data via multi-level modeling as opposed to a more traditional measure, such as the gamma correlation (Nelson, 1984, 1996b). Murayama, Sakaki, Yan, and Smith (2014) found that the use of gamma correlations in a two-stage process to assess group-level differences (namely, by first computing participant-level correlations between JOLs and memory outcomes and then analyzing the derived statistics as dependent variables), can inflate the Type I error rate in the presence of one or more random effects (such as variance in item difficulty and individual differences in the effects of MPT cues on JOLs). Our models estimate fixed effects for each item, which controls for item difficulty effects and directly addresses this concern. Hertzog et al. (2013) found multiple random effects (i.e., individual differences) in analyzing the younger adult data presented here, and our analyses control for these effects, whereas the use of gamma correlations does not. Finally, when marginal distributions of the component variables are highly skewed, gamma correlations have high standard errors that are then ignored in the two-stage process, producing distorted standard errors for aggregate gammas. The multi-level modeling procedure produces appropriate asymptotic standard errors of estimate for all regression effects.

RESULTS

The models presented here focus on key predictors and Trial 2 JOL for both younger and older adults. Relevant means and standard deviations of the variables included therein can be found in Table 1. Briefly, older adults studied longer than did younger adults in order to reach a somewhat dissimilar level of recognition memory performance in Trial 1 (favoring younger adults). Older adults were generally less confident in their memories following both study and test experiences, although, like their younger counterparts, their initial underconfidence in Trial 1 was repaired following test, reflecting the benefits of the MPT heuristic and, possibly, the other potential JOL cues examined in detail below. Of particular note were response time estimates, which were discrepant from actual RT (assessed using the gamma correlations; see Nelson, 1984). Older adults' estimates correlated less with actual RT (M = 0.32, SE = 0.03) than did those of younger adults (M = 0.41, SD = 0.03), d = .46, a medium-sized effect (Cohen, 1988). Given research showing that JOLs are more highly correlated with perceived rather than actual memory fluencies (e.g., Robinson et al., 2006), this discrepancy provides a good opportunity to determine whether our participants (especially older adults) were attuned to actual RT, estimated RT, or both when estimating JOLs.

Multi-level model rationale

We first defined an unconditioned model in which no effects were specified other than the intercept (see Table 2). This model served as a baseline for comparison with our other models because we used it to calculate pseudo-R2 values for each iteration by subtracting the residual (unexplained) variance for those models from the residual of the unconditioned model and dividing that amount by the residual of the unconditioned model. We were thus able to determine the benefit of adding specific components to the model in terms of the additional variance explained in item-level Trial 2 JOL. A second model included a random effect for the intercept, which allowed an examination of the influence of individual differences in average Trial 2 JOL (i.e., memory confidence). A third model added all fixed effects for the potential JOL cues specified above. A fourth model included the aforementioned effects in addition to a random effect of Trial 1 recognition accuracy. As discussed previously, prior research on MPT reported that a large proportion of the variation in subsequent memory confidence (relative to that of other potential cues) is based on whether an individual remembers the item-specific results of a prior memory test, so this effect is hypothesized to be of considerable importance. A reliable fixed effect would indicate that prior response accuracy was an important source of variation in Trial 2 JOL. A reliable random effect would indicate that individuals varied in the extent to which they based JOLs on prior recognition accuracy. A fifth model also included a random effect for Trial 1 CJs, which captured individual differences in this cue's degree of influence on JOLs above and beyond differences in recognition memory accuracy itself. Model 6 retained the random effect of CJs but did not include the random effect of accuracy and served as a test of whether or not individual differences in Trial 1 recognition accuracy had a significant effect on the model's explanatory power if we also included CJs.

Table 2. Goodness-of-fit statistics, LR χ^2 and psuedo-R^2 for models predicting trial 2 JOLs for younger and older adults.

Model	−2LL	BIC	Residual	$\Delta\chi^2$	Δdf	R^2
Younger adults						
1. Unconditioned	28,146	28,163	578.42	—	—	—
2. Random intercept	27,250	27,262	408.81	896.0*	1	.293
3. Add fixed effects	26,439	26,738	315.17	811.7*	0	.455
4. Add random accuracy	26,375	26,682	302.95	63.5*	2	.476
5. Add random CJ	26,325	26,643	295.17	50.7*	3	.490
6. Drop random accuracy	26,343	26,650	299.49	18.3*	3	.482
Older adults						
1. Unconditioned	49,788	49,805	812.45	—	—	—
2. Random intercept	46,780	46,794	427.11	3,007.70*	1	.474
3. Add fixed effects	45,529	45,935	339.71	1,322.08*	0	.582
4. Add random accuracy	45,440	45,855	327.85	89.04*	2	.596
5. Add random CJ	45,352	45,781	319.09	87.23*	3	.607
6. Drop random accuracy	45,379	45,795	322.82	27.00*	3	.603

Notes: R^2 is the total pseudo-R^2 for all effects (fixed and random).
*$P < .001$.
Abbreviations: −2LL: −2 log likelihood (fit function); BIC: Bayesian information criterion; LR: likelihood ratio; JOL: judgment of learning; CJ: confidence judgment.

Multi-level model results

Fit statistics pertaining to our models can be found in the top panel of Table 2. The following description of findings pertains only to older adult models, although the pattern of improvement in model fit resulting from the inclusion of specific random effects was similar between younger and older adults (see Hertzog et al., 2013 for further details regarding younger adult models). Model 2, which only included a random effect on intercept, accounted for approximately 47% of the variance in Trial 2 JOL and offered a better fit to the data than did our initial unconditioned model. Participants therefore differed reliably in their average level of learning confidence as assessed by Trial 2 JOLs. The explanatory power of the model increased considerably following the inclusion of all fixed effects, pseudo-R^2 = .58. Smaller increases were found following the inclusion of random effects of Trial 1 recognition accuracy (pseudo- R^2 = .60) and Trial 1 recognition CJ (pseudo- R^2 = .61). A small but statistically significant reduction in model fit (accounting for 1% of variance explained) resulted from the removal of the random effect of accuracy from the prior model. Individuals therefore varied with respect to the degree of emphasis placed on both recognition accuracy and CJs and the inclusion of both associated random effects in the model is therefore justified. The random effects covariance matrix for the accepted model, Model 5, are reported in Table 3.

Table 3. Covariance matrices of random effects for model 5 (standard errors in parentheses).

	Intercept	Trial 1 Accuracy	Trial 1 CJs
Intercept	159.03 (25.13)***	–	–
Trial 1 accuracy	14.84 (13.91)	37.36 (14.18)**	–
Trial 1 CJs	−0.57 (0.31)	0.29 (0.20)	0.02 (0.01)***

Note: *$P < .05$; **$P < .01$; ***$P < .001$

Table 4. Multi-level regression predicting trial 2 judgments of learning for younger and older adults.

Effect	Estimate	*SE*
Intercept	69.55	3.71
Age (younger adults)	3.88	5.97
Trial 1 study time (within subjects)	0.01	0.06
Trial 1 study time (between subjects)	−0.22	0.41
Trial 1 JOL (within subjects)	0.17	0.03
Trial 1 JOL (between subjects)	0.55	0.14
Trial 1 recognition accuracy (within subjects)	8.55	1.62
Trial 1 recognition accuracy (between subjects)	23.91	13.97
Trial 1 CJ (within subjects)	0.13	0.04
Trial 1 CJ (between subjects)	0.44	0.17
Trial 1 recognition RT (within subjects)	0.06	0.14
Trial 1 recognition RT (between subjects)	3.40	1.80
Trial 1 recognition RT estimate (within subjects)	−0.63	0.47
Trial 1 recognition RT estimate (between subjects)	−1.79	1.64
Trial 2 study time (within subjects)	−0.35	0.07
Trial 2 study time (between subjects)	−1.24	1.11
Trial 1 study time (within subjects) × Age	−0.02	0.08
Trial 1 study time (between subjects) × Age	−0.12	0.59
Trial 1 JOL (within subjects) × Age	−0.07	0.03
Trial 1 JOL (between subjects) × Age	−0.33	0.18
Trial 1 recognition accuracy (within subjects) × Age	3.47	2.19
Trial 1 recognition accuracy (between subjects) × Age	−3.50	25.34
Trial 1 CJ (within subjects) × Age	0.01	0.05
Trial 1 CJ (between subjects) × Age	−0.09	0.32
Trial 1 recognition RT (within subjects) × Age	0.01	0.38
Trial 1 recognition RT (between subjects) × Age	−4.79	4.58
Trial 1 recognition RT estimate (within subjects) × Age	−0.42	0.73
Trial 1 recognition RT estimate (between subjects) × Age	1.46	2.86
Trial 2 study time (within subjects) × Age	−0.31	0.17
Trial 2 study time (between subjects) × Age	2.00	1.40

Note: JOL, judgment of learning; CJ, confidence judgment. Temporal measures are scaled in seconds. Effects associated with specific word-pair cues were not included in this table to conserve space.

The fixed effects solution for Model 5 is reported in Table 4, and the relevant *F*-tests are presented in Table 5. There was a robust main effect for Stimulus, indicating that items varied in their average JOL at Trial 2 (range: 54.31–84.58%). All other effects control for these average between-item differences in perceived degree of learning. The intercept was defined by the average JOL confidence of the older age group, given that age was dummy coded with a zero for older adults and one for younger adults. Therefore, each of the reported age-related effects in

Table 4 can be interpreted as the differential consequence of being a younger adult. With that in mind, the average Trial 2 JOL for older adults was approximately 70%, $t(168) = 18.73$, $P < .001$. The age difference (approximately 4%, favoring younger adults) did not approach statistical reliability ($P = .52$, $d = 0.14$), indicating similar levels of post-study confidence in degree of learning for both age groups in Trial 2. This is useful because any age differences in factors influencing Trial 2 JOLs cannot be attributed to scale sensitivity differences (as might be the case if young and old responses tended to occupy different ranges of the JOL scale).

Table 5. *F*-Tests for multi-level regression fixed effects.

Effect	*df*	*F*	*P*
Word-pair stimulus	5076	2.40	<.001
Age (younger adults)	89.4	0.42	0.517
Trial 1 study time (within subjects)	5063	0.02	0.885
Trial 1 study time (between subjects	89.6	0.93	0.339
Trial 1 JOL (within subjects)	5120	66.73	<.001
Trial 1 JOL (between subjects)	88.9	17.17	<.001
Trial 1 recognition accuracy (within subjects)	73.9	87.62	<.001
Trial 1 recognition accuracy (between subjects)	87.4	3.06	0.084
Trial 1 CJ (within subjects)	89.8	34.45	<.001
Trial 1 CJ (between subjects)	87.3	6.24	0.014
Trial 1 recognition RT (within subjects)	4818	0.12	0.727
Trial 1 recognition RT (between subjects)	88.4	0.19	0.664
Trial 1 recognition RT estimate (within subjects)	4821	5.22	0.022
Trial 1 recognition RT estimate (between subjects)	88.7	0.55	0.462
Trial 2 study time (within subjects)	5069	35.28	<.001
Trial 2 study time (between subjects)	89.1	0.12	0.73
Trial 1 study time (within subjects) × Age	5065	0.09	0.765
Trial 1 study time (between subjects) × Age	89.6	0.04	0.839
Trial 1 JOL (within subjects) × Age	5121	5.19	0.023
Trial 1 JOL (between subjects) × Age	88.9	3.28	0.074
Trial 1 recognition accuracy (within subjects) × Age	73	2.51	0.118
Trial 1 recognition accuracy (between subjects) × Age	87.4	0.02	0.891
Trial 1 CJ (within subjects) × Age	89.1	0.02	0.878
Trial 1 CJ (between subjects) × Age	87.3	0.09	0.767
Trial 1 recognition RT (within subjects) × Age	4835	0.00	0.98
Trial 1 recognition RT (between subjects) × Age	88.4	1.10	0.298
Trial 1 recognition RT estimate (within subjects) × Age	4794	0.33	0.567
Trial 1 recognition RT estimate (between subjects) × Age	88.7	0.26	0.61
Trial 2 study time (within subjects) × Age	5074	3.44	0.064
Trial 2 study time (between subjects) × Age	89.1	2.04	0.156

Note: JOL, judgment of learning; CJ, confidence judgment. Temporal measures are scaled in seconds.

Within-person JOL effects indicate the degree to which items vary in JOLs between Trial 1 and Trial 2. For each percentage increase in an item's Trial 1 JOL, its Trial 2 JOL was increased by approximately 0.17% for older adults and approximately 0.10% for younger adults, $t(5121) = -2.28$, $P < .001$, $d = 0.40$, a medium-sized effect (Cohen, 1988). Given that mean item difficulty is controlled for by the Stimulus factor in the regression equation, this effect suggests that older adults' JOL for each item was more stable across trials than younger adults' item JOL. Individual differences in mean JOLs were also relatively stable across trials. For each percentage increase in a person's average Trial 1 JOL, their average Trial 2 JOL increased by approximately 0.55%, $t(88.9) = 3.97$, $P < .001$, $d = 0.84$. A trend reflecting a smaller effect for younger adults (an increase of .22% in a person's Trial 2 JOL for each percent increase in average Trial 1 JOL), also indicated more stability in post-study confidence for older adults, $P = .07$, $d = 0.06$, although the size of the effect was small (Cohen, 1988). Irrespective of this trend, a person's average confidence in their initial learning was a robust predictor of their poststudy confidence in Trial 2 for all participants.

Despite some degree of stability, there was evidence of within-person and between-person changes in JOLs across the two study–test trials. As predicted by the MPT hypothesis, Trial 1 recognition accuracy had a substantial impact on Trial 2 JOL, such that items with accurate Trial 1 recognition responses were given 8.6% higher Trial 2 JOLs, $t(69.8) = 5.30$, $P < .001$, $d = 1.27$, than items that had not been correctly recognized. This MPT effect did not differ reliably between age groups, $P = .12$, $d = 0.34$.

Several relationships emerged that showed MPT in the recognition memory domain to be broader than just a predictive relationship of memory success to subsequent JOLs. First and foremost, Trial 1 recognition test confidence predicted JOLs independently of recognition test accuracy. For each percent increase in an item's Trial 1 CJ, the item's Trial 2 JOL increased by approximately 0.13% ($t(90.1) = 3.70$, $P < .001$, $d = 0.78$). At the between-person level, each percent increase in a person's average Trial 1 CJ, their average Trial 2 JOL increased by approximately 0.44% ($t(87.7) = 2.56$, $P < .05$, $d = 0.55$). This relationship emerged despite the substantial correlation of CJs with correct recognition responses (CJs for correct recognition responses were 37% higher than CJs for incorrect responses).

Second, reliable random effects for recognition accuracy and CJs were found (see Table 3), indicating reliable individual differences in the relationships between these two variables and JOLs (irrespective of the mean-level correlations and age interactions otherwise reported). The correlation between each predictor and Trial 2 JOL was therefore higher for some participants than for others, which suggests that some participants weighed accuracy cues more heavily than CJs when constructing subsequent JOLs, and vice versa. Although it is possible that some participants increased their emphasis on both recognition accuracy and CJs at the expense of other measured predictors (e.g., actual or estimated RT), we did not detect the presence of any further individual differences (i.e., random effects), which would seem to indicate that this give-and-take process was limited to recognition accuracy and CJs.

Third, the participant-estimated latency of recognition test responses also influenced JOLs. For each perceived 1-s increase in item-level estimated Trial 1 recognition RT, the associated item-level Trial 2 JOL decreased by approximately 0.63%, $t(4819) = -1.33$, $P = .184$, $d = 0.04$. Higher perceived recognition fluency was associated with an increase in subsequent JOL confidence for relevant items; however, unlike the pairing of CJs and objective recognition response accuracy, objective recognition response RT was not also a reliable predictor of subsequent JOLs in our model. A possible explanation for this discrepancy is that the influence of objective RTs on subsequent JOLs may have been mediated by other cues present in the model, thereby decreasing the direct association so much so that it was no longer a reliable predictor, whereas the effect of objective recognition response accuracy is so powerful that it remained in the model even after controlling for other cues with which it is correlated (e.g., CJs; see Hines et al., 2009, for an example of this phenomenon in a similar associative recognition task).

In sum, the regression model indicated that multiple aspects of prior recognition test outcomes influenced Trial 2 JOLs, consistent with a multiple-cue utilization perspective.

Perhaps the most critical evidence for multiple cue utilization involved the relationship of Trial 2 encoding experiences on JOLs, independent of effects arising from Trial 1 recognition test experience. At the within-person level, Trial 2 self-paced study time also influenced Trial 2 JOLs. For each 1-s increase in Trial 2 item-level study, the associated JOL decreased by 0.35%, $t(5021) = -5.26$, $P < .001$, $d = 0.15$.

The most critical effect for evaluating age differences concerned the relation of Trial 2 study time to Trial 2 JOLs. Both younger and older adults exhibited higher confidence for their item memory as they invested more time studying each item at Trial 2, independently of the recognition test influences. However, we could not merely accept the null hypothesis of age equivalence given a statistical trend for an Age × Trial 2 Study Time interaction ($P < .10$). The trend reflects a larger decrease for subsequent JOLs for younger adults (.66%) than for older adults (.35%) for each additional second of study time, $d = 0.09$, 95% confidence interval [−0.02, 0.64]. By Cohen’s (1988) standards, this is at best a small effect, and we conclude that any age difference in the influence of Trial 2 encoding fluency is relatively trivial. To make sure that an age difference at Trial 2 had not been suppressed by including Trial 1 Study Time in the model, we re-ran the model without the Trial 1 variable in the equation. The targeted Age × Trial 2 Study Time interaction was not affected.

Contrary to a resource-limitation hypothesis about age differences, both younger and older adults’ JOLs were influenced by the multiple cues we analyzed, as manifested in a general lack of age-related interactions. Even when trends toward age differences were present, the associated effect sizes were generally quite small. To clarify this point, the mean effect size for all age-related interactions was 0.12 ($SD = .14$), ranging from 0.01 to 0.40.

DISCUSSION

The present study shows that both younger and older adults’ JOLs are influenced by multiple metacognitive cues that can act both simultaneously and independently from oneanother. Despite the fact that multiple cues influence JOLs, we saw little indication that older adults are deficient in multiple-cue utilization, consistent with earlier cued-recall evidence (e.g., Tauber & Rhodes, 2012). One might expect older adults, who typically have been out of a formal educational setting for decades, to generate JOLs differently than our sample of younger adult students, who likely have more recent high-stakes experience in doing so. Instead, older adults showed similar MPT effects, and most critically, their encoding fluency at Trial 2 had an additional effect on JOLs – the evidence for multiple cue utilization.

Regarding encoding fluency, this study shows that previous findings indicating age similarity in encoding fluency effects (Robinson et al., 2006) apparently extend to multitrial JOLs as well. Granted, we observed a trend for age differences in the magnitude of the effect of self-terminated study time on JOLs, but the associated effect size was small, and a reliable relationship of encoding fluency on Trial 2 JOLs was detected in both age groups. It is clear from our data that researchers typically find null effects of age on JOL estimation because younger and older adults use available cues in a similar way to construct JOLs.

The analysis revealed an interesting pattern that underscores the value of using multilevel regression models to study multiple-cue influences on JOLs. Reliable individual differences (random effects) were found in the effects of both recognition accuracy and CJs on Trial 2 JOLs. This finding suggests the intriguing possibility that individuals vary in profiles of reliance (the subjective weight they placed upon) on recognition outcomes and CJs as cues when making JOLs. This finding warrants additional research to try to tease apart the conditions under which people may trade off utilization of these cues. For instance, it could be the case that individuals with higher proportions of strong, accurate recollective experiences in associative recognition attend more to recognition accuracy as a cue, whereas persons susceptible to high-confidence mis-recollections (Shing et al., 2009; Wong et al., 2012) show a greater reliance on CJs over recognition memory accuracy. Given older adults' heightened susceptibility to this type of recognition memory illusion, it would be useful to import JOLs into tasks that manipulate false memory in recognition memory contexts (e.g., Fandakova, Shing, & Lindenberger, 2013; Jacoby & Rhodes, 2006) while attending explicitly to the measurement of both individual differences and person × task interaction effects. It is likely that age differences in retrieval monitoring accuracy could have differential impact on JOLs in tasks that produce high rates of memory illusions. For instance, Daniels, Toth, and Hertzog (2009) found that older adults' JOLs were less predictive of subsequent recollection (remember versus know experiences) in a standard single-trial yes–no recognition task (see also Soderstrom, McCabe, & Rhodes, 2012).

Furthermore, some older adults (e.g., individuals with preclinical levels of mild cognitive impairment) may be more susceptible to these effects (Guerdoux, Dressaire, Martin, Adam, & Brouillet, 2012). Given the lack of age differences in the impact of recognition accuracy and CJs on subsequent JOLs in the present data, despite small differences in mean-level recognition memory accuracy and confidence, these effects appear to be minimal in the current sample. This could in part be a consequence of self-paced study, which may have facilitated more accurate recollective experiences in both age groups.

The relative lack of age differences in how cues influence JOLs in this study fits well with the larger literature suggesting spared monitoring of encoding in old age (Hertzog & Dunlosky, 2011; Hertzog et al., 2010). Given that we have extended this finding to the multiple-cue context of multi-trial learning, it seems that reduced-resources views of older adults' JOL accuracy are not well supported by the present data. However, other explanations are possible that could still be reconciled with a reduced resources view of some aspects of metacognitive monitoring. As suggested by Tauber and Rhodes (2012), it could be the case that JOLs simply do not place strong demands on cognitive resources, being based more on relatively automatic or unconscious influences of cues on confidence (see also Shea et al., 2014). These influences could enable monitoring of multiple cues despite a primary focus of conscious attention on the explicit search for information about the past test as envisioned in the MPT heuristic (Serra & Ariel, 2014). Alternatively, age differences in monitoring accuracy might emerge when individuals are attempting to monitor learning online or in tasks with high concurrent demands on cognitive control. In effect, the measurement context of self-paced JOLs may remove processing demands during judgments that constrain active monitoring or monitoring accuracy in at least some task contexts. In any case, the present findings add to a picture of spared metacognitive monitoring of learning using standard JOL measurement techniques, despite strong age effects on what is being monitored – associative learning itself (e.g., Shing, WerkleBergner, Li, & Lindenberger, 2008; Old & Naveh-Benjamin, 2008).

These results have potential implications for the use of metacognitive tasks in neuropsychological assessment of older adults. Although early work on aging, fluid intelligence, and working memory supported a generic argument that frontal lobe deterioration could be responsible for age-related cognitive decline (e.g., West, 1996), recent work in cognitive neuroscience indicates that differentiating sub-regions in prefrontal cortex is critical for understanding cognitive control, executive function, and metacognition (e.g., Fleming, Huijgen, & Dolan, 2012; Speer, Jacoby, & Braver, 2003). Research has shown that areas of anterior and medial prefrontal cortex, particularly ventromedial prefrontal cortex (vmPFC), are important substrates for supporting JOLs (Do Lam et al., 2012; Kao, Davis, & Gabrieli, 2005) and other metacognitive judgments (e.g., Chua, Schacter, & Sperling, 2009a). Older and younger adults activate similar brain regions when making metacognitive judgments (Chua, Schacter, & Sperling, 2009b), including vmPFC and anterior prefrontal cortex, although the latter may be more important for metacognition about percepts than memory (Fleming, Ryu, Golfinos, & Blackmon, 2014). Structural MRI evidence suggests relative sparing of vmPFC volume, compared to other frontal regions such as dorsolateral prefrontal cortex (e.g., Raz et al., 2005, 2010). Most neuropsychological tests of frontal function, such as the Wisconsin card sorting test, focus on tasks that require working memory and task set maintenance and seem to be more dependent on intact dorsolateral prefrontal regions (e.g., Yuan & Raz, 2014) and are sensitive to age (Carp, Gmeindl, & Reuter-Lorenz, 2010; Salthouse & Davis, 2006; Toepper et al., 2014). Metacognitive judgments are also sensitive to frontal damage as revealed by neuropsychological tests (e.g., Bastin et al., 2012; Pannu, Kaszniak, & Rapcsak, 2005), but often without specificity as to localization of frontal damage. The fact that JOL accuracy is well preserved in old age in a number of associative memory tasks suggests that impairments in JOL accuracy, when they occur, would indicate likely pathology in vmPFC and related frontal network connections that are distinct from the lateral areas important for working memory and the control of memory encoding and retrieval (e.g., Badre & Wagner, 2007; Levy & Wagner, 2013). Future research could consider how task batteries that include metacognitive judgments like JOLs could help elucidate patterns of abnormal function in brain regions in older adults, differentiating consequences of localized frontal pathology from the more benign effects of normal aging on metacognitive monitoring.

ACKNOWLEDGMENTS

Jarrod C. Hines is now at the University of Wisconsin – Eau Claire. We would like to extend thanks to the following personnel for their assistance with subject recruitment and data collection: Teri Boutot, Bethany Geist, Devaki Kumarhia, Colin Malone, Melissa McDonald, and Alisha Monteiro.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This research was supported by a grant from the National Institute on Aging, one of the National Institutes of Health [R01 AG024485].

REFERENCES

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition*, 39, 171–184. doi:10.3758/s13421-010-0002-y

Ariel, R., Hines, J. C., & Hertzog, C. (2014). Test framing generates a stability bias for predictions of learning by causing people to discount their learning beliefs. *Journal of Memory and Language*, 75, 181–198. doi:10.1016/j.jml.2014.06.003

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45, 2883–2901. doi:10.1016/j.neuropsychologia.2007.06.015

Bastin, C., Feyers, D., Souchay, C., Guillaume, B., Pepin, J., Lemaire, C., … Salmon, E. (2012). Frontal and posterior cingulate metabolic impairment in the behavioral variant of frontotemporal dementia with impaired autonoetic consciousness. *Human Brain Mapping*, 33, 1268–1278. doi:10.1002/hbm.21282

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, 28, 610–632. doi:10.1016/0749-596X(89)90016-8

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127, 55–68. doi:10.1037//0096-3445.127.1.55

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory and Cognition*, 41, 897–903. doi:10.3758/s13421-013-0307-8

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444. doi:10.1146/annurev-psych-113011-143823

Carp, J., Gmeindl, L., & Reuter-Lorenz, P. A. (2010). Age differences in the neural representation of working memory revealed by multi-voxel pattern analysis. *Frontiers in Human Neuroscience*, 4, Article 217. doi:10.3389/fnhum.2010.00217

Cerella, J., Onyper, S. V., & Hoyer, W. J. (2006). The associative-memory basis of cognitive skill learning: Adult age differences. *Psychology & Aging*, 21, 483–498. doi:10.1037/0882-7974.21.3.483

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009a). Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, 21, 1751–1765. doi:10.1162/jocn.2009.21123

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009b). Neural basis for recognition confidence in younger and older adults. *Psychology and Aging*, 24, 139–153. doi:10.1037/a0014029

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

Cohn, M., & Moscovitch, M. (2007). Dissociating measures of associative memory: Evidence and theoretical implications. *Journal of Memory and Language*, 57, 437–454. doi:10.1016/j.jml.2007.06.006

Craik, F. I. M., & Hay, J. F. (1999). Aging and judgments of duration: Effects of task complexity and method of estimation. *Perception & Psychophysics*, 61, 549–560. doi:10.3758/BF03211972

Daniels, K., 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

Do Lam, A. T. A., Axmacher, N., Fell, J., Staresina, B. P., Gauggel, S., Wagner, T., … Weis, S. (2012). Monitoring the mind: The neurocognitive correlates of metamemory. *PLoS One*, 7, Article e30009. doi:10.1371/journal.pone.0030009

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory and high confidence errors: A misrecollection account. *Psychology and Aging*, 22, 122–133. doi:10.1037/0882-7974.22.1.122

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, 15, 462–474. doi:10.1037/0882-7974.15.3.462

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage Publications.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multi-level models: A new look at an old issue. *Psychological Methods*, 12, 121–138. doi:10.1037/1082-989X.12.2.121

Fandakova, Y., Shing, Y., & Lindenberger, U. (2013). Differences in binding and monitoring mechanisms contribute to lifespan age differences in false memory. *Developmental Psychology*, 49, 1822–1832. doi:10.1037/a0031361

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, 36, 813–821. doi:10.3758/MC.36.4.813

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 238–244. doi:10.1037/0278-7393.33.1.238

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, 58, 19–34. doi:10.1016/j.jml.2007.03.006

Fleischman, D. A., Wilson, R. S., Gabrieli, J. D. E., Bienias, J. L., & Bennet, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging*, 19, 617–625. doi:10.1037/0882-7974.19.4.617

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience*, 32, 6117–6125. doi:10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain: A Journal of Neurology*, 137, 2811–2822. doi:10.1093/brain/awu221

Guerdoux, E., Dressaire, D., Martin, S., Adam, S., & Brouillet, D. (2012). Habit and recollection in healthy aging, mild cognitive impairment, and Alzheimer’s disease. *Neuropsychology*, 26, 517–533. doi:10.1037/a0028718

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology*, 20, 511–528. doi:10.1037/0894-4105.20.5.511

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults’ self-regulation. *Current Directions in Psychological Science*, 20, 167–173. doi:10.1177/0963721411409026

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 22–34. doi:10.1037/0278-7393.29.1.22

Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of learning are influenced by multiple cues in addition to memory for past test accuracy. *Archives of Scientific Psychology*, 1, 23–32. doi:10.1037/arc0000003

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging*, 17, 209–225. doi:10.1037/0882-7974.17.2.209

Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S., & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact?*The Quarterly Journal of Experimental Psychology*, 62, 155–173. doi:10.1080/17470210701855520

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, 46, 939–948. doi:10.1037/a0019812

Hertzog, C., & Shing, Y. L. (2011). Memory development across the lifespan. In K. Fingerman, C. A.Berg, T. Antonucci, & J. Smith (Eds.), *Handbook of lifespan developmental psychology* (pp. 299–330). New York, NY: Springer.

Hertzog, C., & Touron, D. R. (2011). Age differences in memory retrieval shift: Governed by feeling-of-knowing? *Psychology and Aging*, 26, 647–660. doi:10.1037/a0021875

Hertzog, C., Touron, D. R., & Hines, J. C. (2007). Does a time monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? *Psychology and Aging*, 22, 607–624. doi:10.1037/0882-7974.22.3.607

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, 24, 462–475. doi:10.1037/a0014417

Jacoby, L. L., & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science*, 15, 49–53. doi:10.1111/j.0963-7214.2006.00405.x

Kao, Y., Davis, E. S., & Gabrieli, J. D. E. (2005). Neural correlates of actual and predicted memory formation. *Nature Neuroscience*, 8, 1776–1783. doi:10.1038/nn1595

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, 34, 959–972. doi:10.3758/BF03193244

Levy, B. J., & Wagner, A. D. (2013). Measuring memory reactivation with functional MRI: Implications for psychological theory. *Perspectives on Psychological Science*, 8, 72–78. doi:10.1177/1745691612469031

Light, L. L. (2000). Memory changes in adulthood. In S. H. Qualls & N. Abeles (Eds.), *Psychology and the aging revolution: How we adapt to longer life* (pp. 73–97). Washington, DC: American Psychological Association.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models* (4th ed.). Cary, NC: SAS Institute.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527. doi:10.1037/0033-295X.95.4.492

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1287–1306. doi:10.1037/a0036914

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:10.1037/0278-7393.26.5.1170

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G.Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60053-5

Nelson, T. O. (1996a). Consciousness and metacognition. *American Psychologist*, 51, 102–116. doi:10.1037/0003-066X.51.2.102

Nelson, T. O. (1996b). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology*, 10, 257–260. doi:10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, 23, 104–118. doi:10.1037/0882-7974.23.1.104

Pannu, J. K., Kaszniak, A. W., & Rapcsak, S. Z. (2005). Metamemory for faces following frontal lobe damage. *Journal of the International Neuropsychological Society*, 11, 668–676. doi:10.1017/S1355617705050873

Parks, C. M. (2007). The role of noncriterial recollection in estimating recollection and familiarity. *Journal of Memory and Language*, 57, 81–100. doi:10.1016/j.jml.2007.03.003

Price, J., Hertzog, C., & Dunlosky, J. (2008). Age-related differences in strategy knowledge updating: Blocked testing produces greater improvements in metacognitive accuracy for younger than older adults. *Aging, Neuropsychology, and Cognition*, 15, 601–626. doi:10.1080/13825580801956225

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

Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M., & Lindenberger, U. (2010). Trajectories of brain aging in middle-aged and older adults: Regional and individual differences. *Neuroimage*, 51, 501–511. doi:10.1016/j.neuroimage.2010.03.020

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., … Acker, J. D.(2005). Regional brain changes in aging healthy adults: General trends, individual differences, and modifiers. *Cerebral Cortex*, 15, 1676–1689. doi:10.1093/cercor/bhi044

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625. doi:10.1037/a0013684

Robinson, A. E., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. *Aging, Neuropsychology, and Cognition*, 13, 458–478. doi:10.1080/13825580600572983

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, 13, 140–144. doi:10.1111/j.0963-7214.2004.00293.x

Salthouse, T. A., & Davis, H. P. (2006). Organization of cognitive abilities and neuropsychological variables across the lifespan. *Developmental Review*, 26, 31–54. doi:10.1016/j.dr.2005.09.001

Serra, M. J., & Ariel, R. (2014). People use the memory for past-test heuristic as an explicit cue for judgments of learning. *Memory and Cognition*, 42, 1260–1272. Advance online publication. doi:10.3758/s13421-014-0431-0

Shanks, L. L., & Serra, M. J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin and Review*, 21, 445–453. doi:10.3758/s13423-013-0513-1

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18, 186–193. doi:10.1016/j.tics.2014.01.006

Shing, Y. L., Werkle-Bergner, M., Li, S., & Lindenberger, U. (2008). Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General*, 137, 495–513. doi:10.1037/0096-3445.137.3.495

Shing, Y. L., Werkle-Bergner, M., Li, S.-C., & Lindenberger, U. (2009). Committing memory errors with high confidence: Older adults do but children don't. *Memory*, 17, 169–179. doi:10.1080/09658210802190596

Singer, J. D. (1998). Using SAS proc mixed to fit multi-level models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23, 323–355. doi:10.2307/1165280

Snijders, T. A. B., & Bosker, R. J. (2012). *Multi-level analysis: An introduction to basic and advanced multi-level modelling* (2nd ed.). London: Sage.

Soderstrom, N. C., McCabe, D. P., & Rhodes, M. G. (2012). Older adults predict more recollective experiences than younger adults. *Psychology and Aging*, 27, 1082–1088. doi:10.1037/a0029048

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 155–167. doi:10.3758/CABN.3.3.155

Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' judgments of learning in multitrial learning. *Psychology and Aging*, 27, 474–483. doi:10.1037/a0025246

Toepper, M., Gebhardt, H., Bauer, E., Haberkamp, A., Beblo, T., Gallhofer, B., … Sammer, G. (2014). The impact of age on load-related dorsolateral prefrontal cortex activation. *Frontiers in Aging Neuroscience*, 6, Article 9. doi:10.3389/fnagi.2014.00009

Undorf, M., & Erdfelder, E. (2013). Separation of encoding fluency and item difficulty effects on judgements of learning. *The Quarterly Journal of Experimental Psychology*, 66, 2060–2072. doi:10.1080/17470218.2013.777751

Wechsler, D. (1981). *Wechsler adult intelligence scale – Revised*. New York, NY: Psychological Corporation.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, 120, 272–292. doi:10.1037/0033-2909.120.2.272

Wong, J. T., Cramer, S. J., & Gallo, D. A. (2012). Age-related reduction of the confidence–accuracy relationship in episodic memory: Effects of recollection quality and retrieval monitoring. *Psychology and Aging*, 27, 1053–1065. doi:10.1037/a0027686

Yuan, P., & Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: A meta-analysis of structural neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 180–192. doi:10.1016/j.neubiorev.2014.02.005

Zachary, R. (1986). *Shipley institute of living scale revised manual*. Los Angeles, CA: Western Psychological Services.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 293–357). Mahwah, NJ: Erlbaum.