Journal of Memory and Language 107 (2019) 69–79

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

# How attributes and cues made accessible through monitoring affect self-regulated learning in older and younger adults

Renée DeCaro*, Ayanna K. Thomas

*Department of Psychology, Tufts University, United States*

ARTICLE INFO

*Keywords:*
Metamemory
Memory
Self-regulated learning
Aging

ABSTRACT

Research suggests that cues available at the time of monitoring affect metamemorial control. We investigated how self-regulated learning (i.e., restudy choice) varied as a function of retrieval success and access to target-related partial information in the context of a metamemorial monitoring decision. Young and older adults studied unrelated cue–target pairs and made trial-by-trial monitoring judgments during an initial testing phase. Participants chose a subset of cue–target pairs for restudy either from an array of all cues presented simultaneously (Exp. 1) or on a trial-by-trial basis at the time of test (Exp. 2). Retrieval success prior to restudy was an overwhelming predictor of restudy choice for both young and older adults. Commission errors predicted subsequent restudy only when decisions were made on a trial–by–trial basis. Finally, young and older adults differed in the relationship between successful access to partial information and restudy.

## Introduction

Learners often use tests as a means of assessing current states of learning (Kornell & Son, 2009) and, in doing so, may be able to better predict whether information is likely to be remembered in the future (e.g., Nelson & Dunlosky, 1991). Less well understood are the different types of metamemorial cues available at the time of testing, and how those specific cues could have consequences for self-regulated learning. The present study focused on objective and subjective cues available during testing and subsequent self-regulated learning (SRL) in young and older adults.

Self-regulated learning is a metacognitive process comprised of two interacting processes: monitoring and control (e.g., Nelson & Narens, 1990). Monitoring involves the awareness and assessment of current states of learning and retrieval, whereas control is the behavioral consequence of monitoring. Control can include decisions about what information should be learned and when and how to learn that information (for review, see Dunlosky, Serra, & Baker, 2007). While these processes interact, most research examining the relationship between monitoring and control has investigated how monitoring affects control (but see, Koriat, Ma'ayan, & Nussinson, 2006).

### *Available cues and self-regulated learning*

Monitoring may be assessed in a number of ways. For example, individuals may be asked to provide monitoring assessments prior to a learning phase (e.g., ease-of-learning, Leonesio & Nelson, 1990) or to assess future retrieval of recently studied information (e.g., immediate or delayed judgments of learning (JOLs), Nelson & Dunlosky, 1991). Individuals may be asked to judge future recognizability of unrecalled information (e.g., feeling of knowing (FOK), Hart, 1965) or to provide confidence in the accuracy of retrieved answers (e.g., confidence judgments). In each of these contexts, the act of monitoring makes available cues that learners may use to guide subsequent control of learning or restudy processes. In the present study we focus on cues available during JOLs.

Most researchers assert that successful control of learning depends on the accuracy of monitoring (but see Kimball, Smith, & Muntean, 2012). In paired associate learning, two factors likely determine how accurate (or diagnostic) judgments of learning will be of future retrievability: the information present and the timing of the judgment itself. First, judgments made when both cue and target are present are less accurate than those made when only the cue is present (Dunlosky & Nelson, 1992, 1994), likely because the presence of the target reduces the likelihood of attempted retrieval. Second, cue–only judgments become more accurate after a delay (Dunlosky & Nelson, 1994), because the delay may result in a retrieval attempt that yields more diagnostic cues (Benjamin & Bjork, 1996). In each of these cases, monitoring is accurate to the extent that attempting retrieval is diagnostic of future retrieval.

* Corresponding author at: Department of Psychology, Tufts University, Medford, MA 02155, United States.
*E-mail address:* renee.decaro@tufts.edu (R. DeCaro).

https://doi.org/10.1016/j.jml.2019.04.002
Received 11 October 2018; Received in revised form 29 March 2019
Available online 22 April 2019

Monitoring judgments may not always be based on an attempt to retrieve the target (e.g., Son & Metcalfe, 2005). Monitoring may also be based on the familiarity of or ease–of–processing of the study material (Schwartz, 1994). Indeed, there is evidence to suggest that delayed JOLs may be based on a two–stage process (Metcalfe & Finn, 2008; Son & Metcalfe, 2005), where a pre–retrieval stage determines whether a second, retrieval stage will occur. In a representative study, learners instructed to first attempt retrieval prior to making JOLs demonstrated reaction times that decreased linearly/monotonically with increasing JOL ratings (i.e., reaction times for the lowest rated items were longest). When learners were not instructed to first attempt retrieval, reaction times resembled an inverted U shape function (i.e., reaction times for the lowest and highest rated items were quickest). This pattern suggests that participants report a low JOL without attempting further retrieval, basing the judgment instead on pre–retrieval cues such as the familiarity of the stimulus used to probe memory. Accessible mnemonic cues such as the effort and success of attempting retrieval may be important determinants of subsequent study behavior. However, when monitoring is assessed through JOLs without instructions to first attempt retrieval, it is unclear which cues (e.g., familiarity, retrieval fluency) will be most important.

Attempting retrieval itself may produce useful cues. For example, commission errors made during recall may signal access to semantic features of the unrecalled target (Koriat, Levy-Sadot, Edry, & de Marcas, 2003). Further, when an individual cannot successfully retrieve a target, she may still be able to access partial or contextual information about it (e.g., Brewer, Marsh, Clark-Foos, & Meeks, 2010; Cook, Marsh, & Hicks, 2006; Thomas, Bulevich, & Dubois, 2011). The accuracy of retrieved partial information influences the magnitude and prediction accuracy of monitoring (Brewer et al., 2010; Hertzog, Dunlosky, & Sinclair, 2010; Hertzog, Fulton, Sinclair, & Dunlosky, 2014; Thomas, Bulevich, & Dubois, 2011, 2012). Whether commission errors or target accessibility will be an important source of feedback for self-regulatory control has not been investigated.

In SRL studies, learners can exert control by choosing to restudy a subset of items over others (e.g., Tullis & Benjamin, 2012) or by allocating study time across all items (study time allocation; e.g., Dunlosky & Connor, 1997). In these studies, the relationship between monitoring and the control manipulation that follows is calculated to determine how in line one's assessments (i.e., JOLs) are with control for each individual. One strategy for learners is to select or to spend the longest time studying those items the individual feels are the least well known (e.g., items they've assigned the lowest monitoring ratings). The items for which the individual feels the lowest levels of confidence are those that are subsequently given the most study time or effort, a strategy termed *discrepancy reduction* (Dunlosky & Hertzog, 1998; Nelson & Leonesio, 1988). For example, learners often do not choose to restudy successfully recalled items, instead choosing to restudy those items not successfully recalled (e.g., Karpicke, 2009). Another strategy for learners is to select items that are closer to being known (region of proximal learning, Kornell & Metcalfe, 2006). The items the individual feels are the closest to being mastered are selected or given the most study time, a strategy that is often deployed when study time is limited substantially (Metcalfe, 2002) or when mastery goals are low (e.g., Thiede & Dunlosky, 1999).

Evidence for how commission errors or target accessibility could affect subsequent SRL is, at present, only indirect. For example, learners underestimate the benefits of generating errors in promoting learning as compared to conditions where learners are not required to generate errors (Huelser & Metcalfe, 2012; Yang, Potts, & Shanks, 2017). Expectations about the learning efficacy of errors may lead individuals to adopt different strategies in subsequent SRL. However, in these studies, learners receive immediate corrective feedback following errors. When individuals do not receive corrective feedback, it is less clear if and how errors will affect SRL. Further, when an individual cannot successfully retrieve a target but can successfully access partial or contextual information about it, it is unclear if there will be consequences for SRL.

Successfully retrieving partial information may serve as an indication that the item is closer to being learned, or within a given individual's region of proximal learning. If individuals should be more likely to select these items relative to those for which they could not access partial information, then SRL would be in line with the RPL model. If, however, individuals are more likely to select items for which they have not retrieved partial information, then SRL would be in line with the discrepancy reduction model. Finally, it could be possible that individuals do not use access to partial information as a cue for SRL.

The present study will be the first to examine how control of subsequent study will be influenced concurrently by retrieval, retrieval–based cues, and monitoring. To that end, we will use monitoring that occurs after explicit retrieval, known as the pre-judgment recall and monitoring methodology (PRAM; Nelson, Narens, & Dunlosky, 2004). In a traditional JOL method, participants make only a monitoring judgement. In the PRAM procedure, participants are first asked to attempt target retrieval prior to each monitoring judgment. The benefit of this procedure is that retrieval is explicitly measured, and that the impact of monitoring may be analyzed separately from retrieval success. Specifically, after accounting for retrieval success and retrieval–based cues, we will determine whether monitoring remains relevant for the control decisions one makes when regulating learning.

### *Self-regulated learning and older adults*

Explicit retrieval, concurrent monitoring, and accessible retrieval–based cues may be especially important to examine in populations who demonstrate declines in recollection accuracy, such as older adults. In research examining metacognitive monitoring and control in older and younger adults, age-invariance in monitoring accuracy is sometimes found (see Castel, Middlebrooks, & McGillivray, 2016 for review). However, several studies have found that older adults demonstrate metacognitive control deficits when compared to younger adults. For example, older adults did not allocate study time in accordance with monitoring judgments to the same extent as young adults (Dunlosky & Connor, 1997; Miles & Stine-Morrow, 2004; Price & Murray, 2012; Souchay & Isingrini, 2004, but see Hines, Touron, & Hertzog, 2009). That older adults exhibit a different relationship between monitoring and control than young adults, despite age-invariance in monitoring accuracy, remains an intriguing and unresolved question in the aging and SRL literature.

We suggest that this age-related disconnect may be due in part to older adults' differential dependence on and access to cues available during monitoring. For example, older adults' monitoring may depend on cue–target relatedness to a greater extent than younger adults (Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Thomas, Lee, & Balota, 2013). Evidence also suggests that older adults rely more on, and place greater confidence in, the metacognitive experiences of familiarity and fluency compared to younger adults (Jacoby & Rhodes, 2006). This overreliance and overconfidence may bias subsequent control. For example, reliance on familiarity to guide subsequent control decisions may be misplaced in certain contexts, such as when familiarity is artificially inflated via priming (Hanczakowski, Zawadzka, & Cockcroft-McKay, 2014).

One way to reduce reliance on familiarity and fluency is to require learners to covertly or overtly attempt retrieval prior to engaging in monitoring (Son & Metcalfe, 2005). For older adults, explicit retrieval may be an important predictor of subsequent control (e.g., Dunlosky & Connor, 1997). Moreover, failures to retrieve may reduce overconfidence (Miller & Geraci, 2014). Indeed, when monitoring immediately follows an explicit attempt to retrieve the target (i.e., feeling of knowing monitoring judgments), older adults demonstrate lower absolute confidence in expected future performance than younger adults (Hertzog & Touron, 2011; MacLaverty & Hertzog, 2009; Sacher, Isingrini, & Taconnat, 2013). The differences in reliance on cues of familiarity and differences in magnitude of confidence may have important consequences for subsequent control decisions, and, have thus far remained an understudied topic within the SRL literature in general.

In Experiment 1, young and older adults made restudy selections after the initial study, retrieval, and monitoring phase was completed. During the selection phase, participants could see all possible choices in an array. Presenting all cues at once likely highlights differences among the items. These apparent differences should then promote strategic regulation based on item comparisons (Thiede & Dunlosky, 1999). Prior research indicates that when cues are presented simultaneously during a selection phase, young and older adults can exhibit identical selection behaviors (Dunlosky & Hertzog, 1997; Tullis & Benjamin, 2012). When decisions are made on a trial-by-trial basis, the functions and feedback of monitoring and control alternate. In a sequential presentation, control decisions follow monitoring, but also each control decision can inform or provide feedback for subsequent monitoring (Koriat et al., 2006). In Experiment 2, participants made their selections during the initial testing phase, following each monitoring judgment. That is, participants were given the option to select whether to restudy a given item on a trial-by-trial basis immediately following attempting retrieval and monitoring.

## Experiment 1

In Experiment 1, monitoring occurred after an explicit retrieval attempt. Control was implemented as the decision to restudy a subset of items. We predicted a negative relationship between retrieval success and item selection (selecting a given item for restudy) as well as a negative relationship between JOLs and item selection. We expected that the magnitude of these relationships would differ for young and older adults (Souchay & Isingrini, 2004), and tested whether the relationship between monitoring and control differed as a function of retrieval success and access to retrieval cues.

### *Method*

#### *Participants*

We tested young adults from Tufts University and older adults from the Greater Boston area. Based on prior studies which evaluated metacognitive effects using similar analytic procedures (Hines, Hertzog, & Touron, 2015; Rich, Van Loon, Dunlosky, & Zaragoza, 2017; Sitzman, Rhodes, & Tauber, 2014), we tested 64 participants. Thirty-two undergraduate students (11 males and 21 females), ranging in age from 18 to 22 ($M = 18.78$, $SD = 0.98$), from Tufts University participated for partial fulfillment of class requirements. All young adults tested were fluent English speakers as demonstrated by English equivalency exams prior to entry into the Tufts University undergraduate program. Thirty-two community-dwelling older adults participated in exchange for compensation ($15). Two older adults were excluded on the basis of performance on the initial test (0 correct items). The thirty older adults (4 males and 24 females) included in the analyses ranged in age from 57 to 84 years ($M = 70.93$, $SD = 5.9$). Older adults tested were fluent English speakers with at least 30 years of English as a primary language.

All older adults were administered the National Adult Reading Test (NART; Nelson & Willison, 1991), validated for American older adults (Grober, Sliwinsk, & Korey, 1991). All older adults had an estimated premorbid IQ greater than 100 (which ranged from 103.84 to 130.24). No participant was excluded on the basis of performance on the NART. Older adults had higher mean years of education, $M = 16.47$, $SD = 2.8$, compared to younger adults, $M = 13.08$, $SD = 1.3$, $t(40.1) = 6.00$, $p < .001$, as well as higher scores on the Shipley vocabulary scale (Zachary, 1991), $M = 15.53$, $SD = 2.3$, relative to younger adults, $M = 13.78$, $SD = 1.7$, $t(60) = 3.48$, $p < .01$. Differences in years of education and vocabulary are common in cross sectional studies comparing older and younger adults and were expected to have no significant impact on the evaluation of our hypotheses.

#### *Materials*

The 36 word pairs used in this experiment were from Thomas et al., 2011. In that study, target words were selected from the Affective Norms for English Words word list (ANEW; Bradley & Lang, 1999) and cue words from the University of South Florida Free Association Norms (USF FAN; Nelson, McEvoy, and Schreiber, 2004). Target word valence was used to assess access to partial information. Valence ratings could range from 1 to 9. Negative words used in the experiment ranged in valence ratings from 1.98 to 3.41 ($M_{Neg} = 2.81$, $SD = 0.20$) and positive words ranged in valence ratings from 6.32 to 8.23 ($M_{Pos} = 7.41$, $SD = 0.32$). Note that these ratings are from updated norming data for ANEW words (Warriner, Kuperman, & Brysbaert, 2013). We also evaluated paired word association and word frequency. We checked that no target word occurred in the set of possible associates for its cue, and vice versa (USF FAN; Nelsonet al., 1998). Mean word frequency was 9.56 log HAL ($SD = 1.46$, Range = 6.4–12.3). We obtained this frequency measure from the English Lexicon Project web site (elexicon.wustl.edu; Balota et al., 2007), and note Thomas et al., 2011 used Kučera and Francis (1967) norms. All words are included in Appendix A.

#### *Procedure*

Participants provided informed consent prior to participating in the experiment, which was conducted with E-Prime software (Psychology Software Tools, Pittsburgh, PA) on Windows personal computers. Some differences in testing conditions for the age groups were undertaken. Older adults were tested individually, as responses were spoken out loud and typed by the experimenter. Younger adult participants were tested in groups of one to four, and all responses were typed via keyboard by the participants. Differences in testing between young and older adults is a common practice (e.g., Thomas et al., 2011; Marsh, Dolan, Balota, & Roediger, 2004), as older adults may have less familiarity and experience less comfort with computer-based testing (American Psychological Association, 2014, p. 39). Experimenter entry of typed responses likely influenced responses times; however, comparing response times of age groups was not of interest to the present study. A presentation rate of 4 s per pair for younger adults and 8 s per pair for older adults were selected, as different rates are often used to mitigate age differences in memory performance (e.g., Tullis & Benjamin, 2012). Younger adults were presented with 30 word pairs and older adults were presented with 24 word pairs. All other aspects of the procedure were the same for young and older adults. Although we recognize that group testing for younger adults introduces a confound, practical reasons influenced this decision. In addition, young adults tested individually did not perform differently than young adults tested in the context of a group.

*Initial study and testing phases*. Participants received word pairs in a random order during the initial study phase. Word pairs appeared in the center of the computer screen in black Garamond 32-point font on a white background. Between each word pair presentation, a blank screen appeared for 0.5 s. After studying all word pairs, participants were presented with each cue word individually and were asked to recall the target word (i.e., cued recall). After -each cued recall attempt, participants received two question prompts: one about the valence of the target word and the other a judgment of learning. Participants were asked about the valence of the target word (*Was the second word positive or negative?*) and could respond '*no answer*', '*positive*', or '*negative*' (keys 0, 1, 2 respectively). The JOL prompt was phrased as: *If you can study this item later, how likely do you think it is you can remember the correct answer in the future memory test?* A scale was displayed at the bottom of the screen from 0 to 100 (in increments of 10), where 0 was '*not likely*' and 100 '*very likely*'. After participants had made their responses, the next cue was presented on the screen, and this continued until the participants made a cued recall attempt and answered the questions for every studied cue.

*Restudy and final test.* After finishing the initial testing/monitoring phase, participants were instructed to select half of the items to restudy. The entire set of cues was presented on the screen in an array comprised of 6 columns (Thiede & Dunlosky, 1999). Four array orders were presented across participants, so that each cue word appeared an equal amount of times in the four different quadrants of the computer screen. Each cue was labeled with a number, and items were selected by typing the cue's corresponding number. The cues disappeared from the array when they were selected, and the number of items that remained to be selected was updated to display the selections remaining. When all allotted cues had been selected, participants began the restudy phase. During the restudy phase, the selected word pairs were re-presented similarly to the first study phase, but in a new random order. Finally, participants engaged in the cued recall task that included all word pairs, where they were given each cue and asked to provide the appropriate target item. The cue remained on the screen through the response, and items were tested in a new random order. Only the initial study and restudy phases were presented at a fixed rate; all remaining phases of the experiment (cued recall, valence question, JOL, and the selection phase) were self-paced.

### *Results and Discussion*

To start, we present descriptive statistics from the initial testing phase. These included the responses to the recall, valence, and JOL prompts. Next, we included resolution measures to address the predictive accuracy of JOLs and their relationship to restudying. To address our main hypotheses, we evaluated cues available during the initial testing phase simultaneously via generalized multilevel model to predict the likelihood of restudy. Due to a programming error, some JOL data points were missing (1.6% in Experiment 1 and 1.1% in Experiment 2). We believe that the probability of the missing data was completely random, and, the trials themselves were omitted (i.e., the number of trials could vary across participants depending on whether they had complete data on JOLs) in analyses with JOLs. For some of the following analyses, listwise deletion occurred when summary statistics (i.e., within-person correlations) could not be computed, and the degrees of freedom were changed accordingly. In cases where the assumption of sphericity was not met, Greenhouse-Geisser correction was applied and the adjusted degrees of freedom were reported to the tenths place.

#### *Initial testing phase descriptive statistics*

We calculated the mean proportion of items correctly recalled on the initial test and the mean proportion of the two types of errors (commission and omission). Commission errors were any non-blank response which did not identically match the target word. Example errors include those that shared semantic features of the target (e.g., "funny" for the target "joke"), shared phonetic features (e.g., "long" for the target "loss"), shared both (e.g., "sunlight" for "sunshine"), were targets of another word pair, or were none of the above. As can be seen in Table 1, the greatest proportion of responses for both young and older adults were errors of omission (i.e., no response to the recall prompt). Important for the present study, young and older adults did not differ in the proportion of words correctly recalled on the initial test, $t(60) = 1.06$, $p = .30$.

We also tabulated access to partial information for all items not correctly recalled on the initial test. Participants accessed correct partial information when the valence of their response (e.g., positive) matched that of the target word (e.g., positive). Participants accessed incorrect partial information when the valence of their response (e.g., positive) did not match that of the target word (e.g., negative). Young and older adults demonstrated similar rates of correct access to partial information, $t(56) = 0.23$, $p = .82$. The results highlighted in the present section are important to evaluate in the context of self-regulated learning. Young and older adults demonstrated similar rates of correct responses. Thus, any difference observed in subsequent SRL is likely not attributable to deficits in recollection (e.g., access to partial information), which are commonly present in aging. See Supplement for analyses of commission errors and incorrect access to partial information.

Young and older adults differed in average JOLs, $t(49.3) = 2.51$, $p < .05$. Consistent with previous research, young adults demonstrated greater average JOLs than older adults, $M_{Young} = 58.36$, $SD = 14.17$, $M_{Older} = 46.59$, $SD = 21.79$. We also present descriptive statistics for average JOLs as a function of response type on the initial test in Table 1 and refer the reader to Supplement for accompanying inferential statistics. Further, the distribution of judgments of learning by response on the initial test are displayed in Fig. 1 for illustrative purposes.

Of interest to the present study was whether participants could distinguish between answers they provided that were correct as compared to incorrect. If individuals believed that all answered items (i.e., commission errors and correct responses) were correct, then the belief in the item's future recallability would likely be equal. If, however, individuals were able to monitor the correctness of their response, they should demonstrate monitoring differences between correct and incorrect provided responses. Correct and incorrect answers were subjected to paired sample t-tests. Regarding initial test responses, commission errors differed in average JOL from correct responses, $t(57) = 10.50$, $p < .001$. When individuals provided responses on the initial test which were correct, average JOL was greater than when individuals provided responses that were incorrect (commission errors). Similarly, when individuals accessed partial information (i.e., reported that the target was positive or negative), correct identifications of target valence differed in average JOL from incorrect identifications, $t(44) = 3.72$, $p < .01$. When individuals accessed correct partial information, average JOL was greater than when individuals accessed incorrect partial information. As has been previously demonstrated in the FOK literature (e.g., Thomas et al., 2011), the quality of retrieved partial information (i.e., correctness) influenced the magnitude of monitoring. Age comparisons and complete inferential analyses are reported in the Supplement.

#### *Resolution of judgments of learning*

Our next goal was to evaluate the predictive accuracy of JOLs. We computed Somer's D, an asymmetric extension of the Goodman–Kruskal Gamma correlation, between JOLs and final test performance (the dependent measure) for items studied only once. Average Somer's D was positive for both young and older adults. Individuals were accurate at predicting final test performance. No age differences were present, $t(49) = 0.36$, $p = .72$. Mean Somer's D correlations (as well as Gamma correlations) are presented in Table 2.

We next evaluated the relationship between JOLs and restudy selection, which we also computed via Somer's D. The average negative relationships were different for young ($M = -0.32$ $SD = 0.18$) and older adults ($M = -0.17$ $SD = 0.35$), $t(42.6) = 2.10$, $p < .05$, $d = 0.54$. The relationships reliably differed from zero for both young adults, $t(31) = 10.16$, $p < .001$, and older adults, $t(29) = 2.69$, $p < .05$. Both age groups were more likely to select items they assigned lower JOLs. Consistent with previous research, older adults did not regulate learning in accordance with monitoring judgments to the same extent as younger adults (Souchay & Isingrini, 2004). Monitoring resolution of other predictors (e.g., the relationship between partial information and restudy selection) is included in the Supplemental analyses.

Results from this section demonstrated that while young and older adults were similarly accurate at predicting final test performance from JOLs, they differed in the magnitude of the relationship between JOLs and subsequent restudy. While illustrative, the average relationships reported above do not account for multiple predictors. It is possible that the relationship between JOLs and restudying varies as a function of other predictors, such as response type on the initial test. Therefore, we investigated restudy selection via mixed-effects modeling, to

**Table 1**
Mean rates of responses and average JOLs by response type by age group. SDs in parentheses.

| | Experiment 1 | | | | | |
|---|---|---|---|---|---|---|
| | Young Adults | | | Older Adults | | |
| **All Test Responses** | Correct | Incorrect | None | Correct | Incorrect | None |
| Initial Test Response Rates | 0.33 (*0.16*) | 0.34 (*0.23*) | 0.46 (*0.25*) | 0.28 (*0.22*) | 0.22 (*0.24*) | 0.53 (*0.30*) |
| Initial Test and JOLs | 83.6 (*13.0*) | 55.0 (*18.1*) | 40.9 (*19.4*) | 70.0 (*18.7*) | 49.6 (*21.9*) | 35.7 (*22.1*) |
| **Incorrect Responses** | | | | | | |
| Partial Information Rates | 0.39 (*0.23*) | 0.18 (*0.13*) | 0.53 (*0.28*) | 0.40 (*0.21*) | 0.21 (*0.17*) | 0.54 (*0.30*) |
| Partial Information and JOLs | 56.2 (*15.8*) | 45.2 (*23.4*) | 35.8 (*20.0*) | 46.3 (*19.9*) | 40.8 (*17.8*) | 32.9 (*25.0*) |
| | Experiment 2 | | | | | |
| | Young Adults | | | Older Adults | | |
| **All Test Responses** | Correct | Incorrect | None | Correct | Incorrect | None |
| Initial Test Response Rates | 0.34 (*0.24*) | 0.46 (*0.28*) | 0.34 (*0.24*) | 0.26 (*0.15*) | 0.22 (*0.14*) | 0.54 (*0.19*) |
| Initial Test and JOLs | 83.4 (*12.5*) | 52.7 (*19.7*) | 43.5 (*16.9*) | 70.4 (*21.5*) | 49.8 (*19.9*) | 36.6 (*17.5*) |
| **Incorrect Responses** | | | | | | |
| Partial Information Rates | 0.36 (*0.21*) | 0.20 (*0.12*) | 0.51 (*0.26*) | 0.40 (*0.20*) | 0.23 (*0.14*) | 0.58 (*0.27*) |
| Partial Information and JOLs | 60.1 (*20.0*) | 51.3 (*20.5*) | 41.5 (*21.3*) | 48.2 (*19.5*) | 44.8(*14.6*) | 31.6 (*17.7*) |

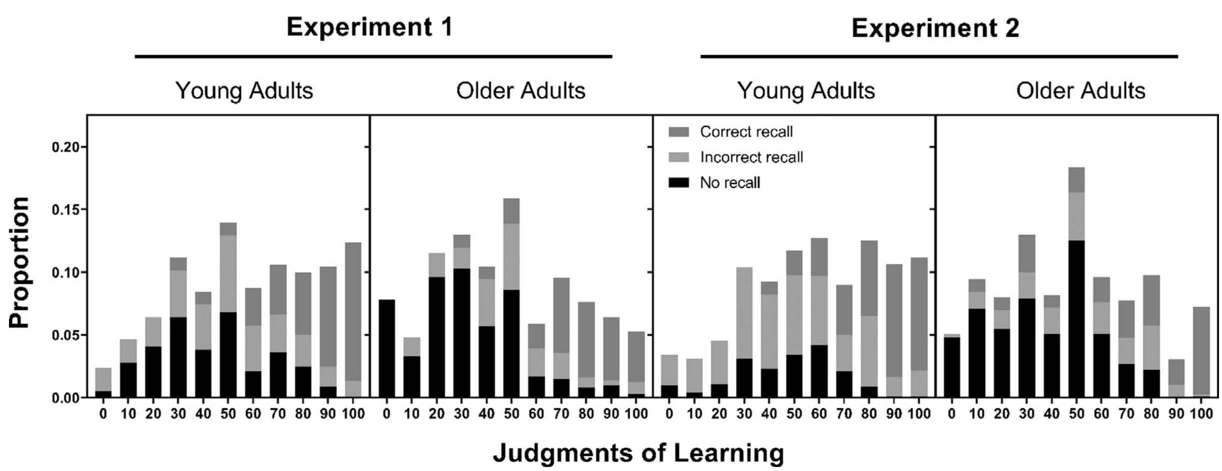


**Fig. 1.** Frequency distribution of proportion of responses at each JOL by initial test response type.

**Table 2**
Final test and predictive accuracy by age group as a function of restudying.

| | | Experiment 1 | | Experiment 2 | |
|---|---|---|---|---|---|
| | | Younger | Older | Younger | Older |
| **Final Test Accuracy** | | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) |
| Items studied once | | 0.55 (*0.25*) | 0.34 (*0.33*) | 0.56 (*0.28*) | 0.35 (*0.27*) |
| Items restudied | | 0.72 (*0.18*) | 0.67 (*0.26*) | 0.74 (*0.19*) | 0.63 (*0.17*) |
| **Prediction Accuracy** | | | | | |
| Items studied once | Somer's D | 0.41 (*0.18*) | 0.42 (*0.14*) | 0.41 (*0.22*) | 0.34 (*0.25*) |
| | Gamma | 0.82 (*0.20*) | 0.88 (*0.20*) | 0.89 (*0.19*) | 0.72 (*0.49*) |

simultaneously evaluate all cues available during the initial test.[1]

*Mixed effects model predicting restudy selection*

We conducted a generalized multilevel model to predict the likelihood of selecting an item to restudy, using R (R Core Team, 2013). We modeled the likelihood of selection with predictors at two levels. At the lowest level (Level 1) were trials; predictors at this level captured variability *within* subjects. At the highest level (Level 2) were subjects; predictors at this level captured variability *between* subjects. Trial level predictors included response type, access to partial information, and JOL. JOL was scaled into units of 10% (e.g., a value of '1′ represents 10%) and centered within subjects (for every trial of a given subject, that subject's mean JOL was subtracted from their trial JOL). The subject level predictor was age group. Response type, access to partial information, and age group were dummy-coded such that no response,

[1] Some differences in these analyses are of note. First, in the mixed-effects analysis, JOLs reflect absolute values as compared to the ordinal analyses reported (i.e., Somer's D) in which differences in JOLs are relative. Second, mixed-effects analyses include random effects associated with participants and items as well as fixed effects (Baayen, Davidson, & Bates, 2008), and can mitigate limitations in the use of by-participant analysis (e.g., Murayama, Sakaki, Yan, & Smith, 2014).

**Table 3**
Summary of generalized multilevel regression predicting likelihood of restudy in Experiment 1.

| Fixed Effects | | β (SE) | z | p |
|---|---|---|---|---|
| Intercept | | 0.53 (*0.18*) | 2.88 | < 0.01 |
| Age (older) | | −0.48 (*0.23*) | 2.06 | < 0.05 |
| $JOL_{within}$ | | −0.08 (*0.09*) | 0.92 | 0.36 |
| $JOL_{within}$ × Age (older) | | −0.01 (*0.13*) | 0.10 | 0.92 |
| Correct partial information | | 0.15 (*0.21*) | 0.70 | 0.48 |
| Correct partial × Age (older) | | 0.60 (*0.30*) | 2.03 | < 0.05 |
| Correct recall | | −1.23 (*0.42*) | 2.89 | < 0.01 |
| Correct recall × Age (older) | | 0.24 (*0.60*) | 0.39 | 0.69 |
| Correct recall × $JOL_{within}$ | | −0.57 (*0.16*) | 3.57 | < 0.001 |
| Correct recall × $JOL_{within}$ × Age (older) | | 0.42 (*0.22*) | 1.96 | 0.05 |
| Incorrect recall | | −0.19 (*0.23*) | 0.83 | 0.41 |
| Incorrect recall × Age (older) | | −0.11 (*0.32*) | 0.34 | 0.73 |
| Incorrect recall × $JOL_{within}$ | | −0.30 (*0.11*) | 2.80 | < 0.01 |
| Incorrect recall × $JOL_{within}$ × Age (older) | | 0.29 (*0.19*) | 1.35 | 0.18 |
| **Random Variance (*SD*)** | | | | |
| Intercept | Item | 0.06 (*0.25*) | | |
| Slope | Correct recall | 1.32 (*1.15*) | | |

no access to partial information, and younger adults were the referents. In sum, for the Level 1 model we added the following fixed effects: JOL, correct partial information, incorrect partial information, correct recall, and incorrect recall (commission errors). Interaction terms were included between recall predictors and JOL to test whether the relationship between JOLs and restudy differed across item response types. For the Level 2 model (final model), we added age as a fixed effect. This included interactions between age and variables at Levels 1. These *cross-level* interactions addressed whether the relationship between a Level 1 predictor (i.e., within-subject JOL) and the outcome differed as a function of some characteristic of the Level 2 grouping variable (in this case, the participant's age group). Our model fitting procedure, including non-significant effects which were trimmed from the model, and likelihood ratio tests are reported in the Supplement. The included variables and parameter estimates of our final model may be found in Table 3.

*Response type.* When participants correctly recalled the target on the initial test, they were less likely to select it to be restudied, relative to test items that received no response. For young adults (the referent), the log-odds of selection were reduced by 1.23; that is, the odds of selection for a correctly recalled item were 0.29 that of the odds of items that received no response. Interestingly, the likelihood of selecting an item to be restudied did not differ between items that received no response and those that were answered incorrectly (commission errors). Notably, the relationship between response type and restudying did not differ between young and older adults.

*Partial information.* We examined whether access to partial information was related to restudy selection. For younger adults (the referent), access to correct partial information did not significantly predict restudy. A significant age by correct partial information interaction was present. For older adults, access to correct partial information was associated with an increase in log-odds of selection of 0.60, relative to younger adults (the referent). When older adults accessed correct partial information, the odds of selection were over twice that of items with no access to partial information, $\exp(+0.15 + 0.60) = 2.12$. Accessing incorrect partial information was not related to the likelihood of selecting an item to be restudied and was trimmed from the model during our model fitting procedure (reported in Supplement).

*Judgments of learning.* We tested whether the relationship between JOLs and restudy differed across item response types. As one can see from Table 3, trial JOL significantly predicted restudy for correctly recalled and incorrectly recalled items but not for unrecalled items. Note that these associated parameter estimates are for young adults (i.e., when age group = 0). For correctly recalled items, for two items that differed in one unit of JOL (10%), the odds of selection for the higher JOL item would be 0.57 that of the odds of selecting an item with one unit lower in JOL. For incorrectly recalled items, the odds of selecting the item with one unit higher JOL would be 0.74, ($\exp(-0.30) = 0.74$). For illustrative purposes we refer the reader to Fig. 1., where the frequency with which individuals assigned different ratings of JOLs is plotted by whether the items were unrecalled or correctly or incorrectly recalled. For correctly and incorrectly recalled items, the likelihood of selecting the items to be restudied decreased as values of JOLs increased. For unrecalled items, individuals were equally likely to select an item to be restudied across values of JOLs.

Older adults did not reliably differ from younger adults. We note, however, that the relationship between JOLs and restudying for recalled items was slightly weaker for older adults. For correctly recalled items, the odds of older adults selecting the higher JOL were 0.86 that of the odds of selecting an item with a lower JOL, $\exp(-0.57 + 0.42) = 0.86$. These odds are closer to 1, reflecting a weaker relationship between JOL and restudying for correctly recalled items. The interaction was marginally significant, $p = .0503$.

Results from the mixed-effects model depicted a more nuanced relationship between predictors and restudy selection. First, restudy selection behavior depended on the *success* of the initial response, and not the response type per se. Individuals were equally likely to select items that were unanswered (i.e., unrecalled) and answered incorrectly on the initial test. Second, judgments of learning were related to restudy selection, but only for items that received a response on the initial test. In both cases, individuals were more likely to select items that were assigned lower JOL ratings. JOLs were not related to restudy selection for unrecalled items. Finally, the mixed-effects results demonstrated, for the first time, that partial information was related to restudy selection. Older adults were more likely to select items when they accessed correct information about the target's valence.

## Experiment 2

In Experiment 1, we found similarities in self-regulated learning for young and older adults. Namely, results from the mixed-effects analysis demonstrated that young and older adults similarly used initial recall success to guide subsequent restudy selection. Moreover, the effect of JOLs on restudy selection depended on the success of prior retrieval. In Experiment 2, we investigated whether these similarities would persist when control decisions were made on a trial-by-trial basis. In contrast to the first study, where participants selected items to restudy only after all items were initially tested, participants in this experiment chose whether to restudy items after each recall attempt.

### Method

#### Participants

Young and older adults from the same participant pool were included in the study; no young or older adults previously participated in Experiment 1. Thirty-two undergraduate students (15 males), ranging in age from 18 to 21 ($M = 18.63$, $SD = 0.8$), from Tufts University participated for partial fulfillment of class requirements. Thirty-two community-dwelling older adults participated in exchange for compensation ($15). Four older adults were excluded on the basis of initial test performance. Twenty-eight older adults (6 males), ranging in age from 62 to 93 years ($M = 74.07$, $SD = 9.4$), were included. As in Experiment 1, older adults were administered the NART, and none were excluded on this basis. Each had an estimated premorbid IQ greater than 100 (which ranged from 108.24 to 127.52). Older adults had higher mean years of education, $M = 17.07$, $SD = 3.3$, compared to younger adults, $M = 12.56$, $SD = 0.8$, $t(30.1) = 7.49$, $p < .001$, as well as higher scores on the Shipley vocabulary scale (Zachary, 1991), $M = 15.78$, $SD = 2.7$, relative to younger adults, $M = 13.97$, $SD = 1.7$, $t(42.6) = 3.00$, $p < .01$.

*Procedure*

Most aspects of the procedure were kept consistent with Experiment 1. Participants first completed the initial study phase. At the time of the initial test, restudy selections were made on a trial-by-trial basis. After participants were asked to provide a JOL, a new screen appeared which displayed only the cue and asked, *"Would you like to restudy this word pair?"*, to which the participant could respond '*Yes*' or '*No*' (keys 1, 2 respectively). Participants were limited to choose a subset of the items to restudy (exactly half), as has been done in prior studies (e.g., Dunlosky & Hertzog, 1997). Two counters appeared on the screen: one that indicated the trial number (e.g., *"Trial 2 of 24"*) and one that indicated how many choices the participant had left. Once a participant had selected the allotted number of choices, the restudy prompt no longer appeared. After the initial testing phase was complete, participants proceeded to the restudy phase.

*Results and Discussion*

We present descriptive statistics from the initial testing phase, measures of resolution, and the results of the mixed-effects analysis. As in Experiment 1, we refer the reader to the Supplement for corresponding inferential statistics.

*Initial testing phase descriptive statistics*

The greatest proportion of responses for older adults were errors of omission (i.e., no response to the recall prompt), while young adults made more errors of commission. As in Experiment 1, we compared young and older adults in the proportion of words correctly recalled on the initial test. Young adults correctly recalled a greater proportion of words than older adults, $t(58) = 2.04$, $p = .046$, 95% CI [0.001, 0.17], $d = 0.53$.[2] When items were not answered correctly on the initial test, young and older adults demonstrated similar rates of correct access to partial information, $t(55) = 0.71$, $p = .48$.

Consistent with Experiment 1, young adults demonstrated greater average JOLs than older adults, $M_{\text{Young}} = 59.78$, $SD = 17.48$, $M_{\text{Older}} = 48.04$, $SD = 18.05$, $t(58) = 2.56$, $p < .05$.

We also compared JOLs as a function of response type on the initial test (see Table 1 for means). When individuals provided responses on the initial test, correct answers were given higher average JOL ratings than were answers that were incorrect (i.e., commission errors), $t(57) = 11.54$, $p < .001$. When individuals accessed partial information, correctly retrieved information was rated with higher average JOLs than was retrieved incorrect information, $t(43) = 3.66$, $p < .01$. In sum, differences in the rates of correct responses were minimal between young and older adults, and average JOLs differed as a function of response type.

*Resolution of judgments of learning*

No age differences were found when we compared average Somers' D calculated between JOL and final test performance for items studied only once, $t(53) = 1.02$, $p = .31$. Means are reported in Table 2. When we examined the relationship between JOLs and restudying, the relationship was negative. Both age groups were more likely to select items given lower JOL ratings ($M_{\text{Young}} = -0.27$, $SD_{\text{Young}} = 0.26$; $M_{\text{Older}} = -0.15$, $SD_{\text{Older}} = 0.37$). The relationships were different than zero for both young adults, $t(30) = 5.86$, $p < .001$, and older adults, $t(26) = 2.10$, $p < .05$, and did not differ from one another, $t(56) = 1.48$, $p = .15$. These results are in contrast to Experiment 1, as we found that young and older adults in Experiment 2 did not differ in the magnitude of the relationship between JOLs and subsequent restudy.

*Mixed effects model predicting restudy selection*

We conducted a generalized multilevel regression to predict the likelihood of selecting an item to be restudied. The same model procedure was followed for Experiment 2 with one exception: cases for which the restudy prompt was never asked were omitted (these cases could never have been selected so the likelihood of selection was known, P(Y) = 0). The included variables and estimates of fixed and random effects for the final model are included in Table 4.

*Response type.* When participants correctly recalled the target on the initial test, they were less likely to select it to be restudied, relative to test items that received no response. For young adults (the referent), the log-odds of selection were reduced by 2.38; that is, the odds of selection for a correctly recalled item were 0.09 that of the odds of items that received no response. Unlike Experiment 1, the likelihood of selecting an item to be restudied differed between items that received no response and those that were answered incorrectly (incorrect recall). In this instance, the odds of selection for an incorrectly recalled item would be 0.39 that of the odds of item that received no response. No age differences were present.

*Partial information.* As in Experiment 1, accessing incorrect partial information was not related to selecting an item to be restudied (see Supplement). Unlike Experiment 1, young adults did demonstrate a relationship between accessing correct partial information and selecting an item to be restudied. Access to correct partial information was associated with a decrease of 0.85 in log-odds of selection. When younger adults accessed correct partial information, the odds of selection were 0.43 that of items with no access to partial information, $\exp(-0.85) = 0.43$. The age interaction was significant; the relationship between accessing correct partial information and restudying was in the opposite direction for older adults. When older adults accessed correct partial information, the odds of selection were over twice that of items with no access to partial information, $\exp(-0.85 + 1.57) = 2.05$. Thus, when young adults accessed correct partial information, their odds decreased, whereas for older adults, their odds increased.

*Judgments of learning.* Trial JOL significantly predicted restudy for correctly recalled items only. For two items that differed in one unit of JOL (10%), the odds of selection for the higher JOL would be 0.35 that of the odds of selecting an item with one unit lower in JOL. No age differences were present.

As in Experiment 1, results from the mixed-effects model depicted a more nuanced relationship between predictors and restudy selection. First, restudy selection behavior depended on the initial response type, as both correctly answered items and commission errors demonstrated a reduced likelihood of selection compared to unanswered items. Second, judgments of learning predicted restudy selection, but only for correctly answered items. Of the items answered correctly on the initial test, individuals were more likely to select items that were assigned lower JOL ratings. Finally, age differences were again apparent in the relationship between partial information and restudying. Older adults were more likely to select items when they accessed correct information about the target's valence, whereas younger adults were less likely to select items when they accessed correct information.

[2] We reran the mixed effects analysis predicting restudy using a subset of 23 young and 23 older adults matched on initial test performance, $t(44) = 0.06$, $p = .95$. All significant effects remained significant.

**Table 4**
Summary of generalized multilevel regression predicting likelihood of restudy in Experiment 2.

| Fixed Effects | | β (*SE*) | *z* | *p* |
|---|---|---|---|---|
| Intercept | | 2.25 (*0.40*) | 5.91 | < 0.001 |
| Age (older) | | −0.46 (*0.47*) | 0.99 | 0.32 |
| $JOL_{within}$ | | 0.07 (*0.15*) | 0.46 | 0.64 |
| $JOL_{within}$ × Age (older) | | 0.16 (*0.18*) | 0.89 | 0.37 |
| Correct partial information | | −0.85 (*0.25*) | 3.45 | < 0.001 |
| Correct partial × Age (older) | | 1.57 (*0.42*) | 3.75 | < 0.001 |
| Correct recall | | −2.38 (*0.68*) | 3.52 | < 0.001 |
| Correct recall × Age (older) | | −0.04 (*0.94*) | 0.04 | 0.97 |
| Correct recall × $JOL_{within}$ | | −1.06 (*0.25*) | 4.24 | < 0.001 |
| Correct recall × $JOL_{within}$ × Age (older) | | −0.47 (*0.36*) | 1.32 | 0.19 |
| Incorrect recall | | −0.93 (*0.40*) | 2.31 | < 0.025 |
| Incorrect recall × Age (older) | | −0.63 (*0.54*) | 1.18 | 0.24 |
| Incorrect recall × $JOL_{within}$ | | −0.16 (*0.16*) | 1.00 | 0.32 |
| Incorrect recall × $JOL_{within}$ × Age (older) | | −0.37 (*0.24*) | 1.55 | 0.12 |
| **Random Variance (*SD*)** | | | | |
| Intercept | Subject | 0.48 (*0.69*) | | |
| Slope | Correct recall | 4.91 (*2.22*) | | |

## General Discussion

Although research has consistently demonstrated that learners may attempt to retrieve information to determine how well that information has been learned, research examining how retrieval-based cues influence subsequent learning strategies has been limited. The present study addressed this important gap in the metacognitive literature and provides insight into how both younger and older adults use information accrued from a retrieval attempt to guide subsequent restudy decisions. In two experiments, we examined how control of study would be influenced by a retrieval attempt made during monitoring, the cues that became available during the monitoring phase, and the monitoring judgment itself.

Results from Experiments 1 and 2 were illustrative in several ways. First, the relationship between JOLs and subsequent restudy decisions depended on the outcome of the initial retrieval attempt. Second, we demonstrated a relationship between access to correct partial information and subsequent restudy. Third, that relationship was directionally different for young and older adults. For older adults, accessing partial information increased the likelihood of restudy selection. For younger adults, accessing partial information was either unrelated to restudy selection or negatively related to restudy selection, depending on the timing and presentation of the selection decision. Finally, across the two experiments, individuals relied on various cues when control decisions were made during a separate control phase, and when control decisions were made in a sequential manner, during the initial testing phase.

### *Accessible cues and restudy selection*

The present results provide support for two complementary models of self-regulated learning. Consistent with the Discrepancy Reduction Model (Dunlosky & Hertzog, 1998; Nelson & Leonesio, 1988), successful retrieval of the target on the initial test was associated with a decreased likelihood of selecting that item to be restudied, in line with results from previous studies (e.g., Dunlosky & Connor, 1997). We also demonstrated that commission errors made on the initial test were associated with a decreased likelihood of restudy selection. With correct responses and errors of commission, individuals can use the response they've produced as a metacognitive cue (Narens, Jameson, & Lee, 1994). Successful recall is a cue used to determine that an item has been sufficiently learned and does not need to be studied again (Karpicke, 2009). Commission errors can signal access to semantic features of the target answer (Koriat et al., 2003), and can indicate a greater degree of underlying learning. In the present study, individuals were more likely to select items for which they could not produce an answer, relative to successfully recalled items (Experiments 1 and 2) and commission errors (Experiment 2). Selecting items that are farther from being known is a strategy consistent with the discrepancy reduction model.

Building upon prior research investigating the role of partial information in monitoring, this study is the first to demonstrate that accessing correct partial information in the presence of inaccurate responses was predictive of selection behavior. Consistent with the Discrepancy Reduction Model, young adults in Experiment 2 were less likely to select items for which they could access correct partial information. However, consistent with the Region of Proximal Learning Model (RPL; Metcalfe, 2002), accessing correct partial information was associated with an increased likelihood of restudy selection for older adults. That is, accessing semantic information of the unrecalled target may indicate that an item is closer to being learned compared to not successfully accessing such information. Older adults' restudy selection strategy seemed to be guided by their evaluation of how likely restudy would result in later successful recall. The RPL model posits that the optimal strategy for a given individual is to study the subset of items closest to one's level of current mastery. Notably, this strategy was employed by older adults in Experiment 1 and Experiment 2, and this strategy was not employed by younger adults in either experiment. Future studies could explore whether this pattern is observed when accessed information is not semantically related to the target (e.g., font color, Schwartz, Pillot, & Bacon, 2014).

### *Retrieval, monitoring, and restudy selection*

Overt retrieval in the context of the present study was illustrative. Across experiments, monitoring reflected the accuracy of responses on the initial test. By using the PRAM methodology (Nelson et al., 2004) and measuring access to partial target information, we were able to measure monitoring in the presence of other relevant, but often overlooked, response markers (e.g., errors of commission). Our results indicated that individuals were able to accurately monitor the correctness of their response on the initial test and were also able to monitor the accuracy of partial information. Specifically, individuals felt least confident in the likelihood of future recallability when they provided no answer on the initial test. When individuals provided an incorrect answer (either a commission error or incorrectly identifying the valence of the target word), average JOL was reliably lower than when a correct answer was provided but higher than when no information was recalled (comparisons reported in the Supplement). The results of the present experiments are consistent with those found in the feeling of knowing literature and suggest that monitoring is based on the partial information retrieved (Koriat, 1993).

Measuring objective performance enabled us to test whether the relationship between monitoring and control differed as a function of retrieval success. In the present study, JOLs were predictive of restudy only when individuals produced an answer on the initial test. Results from Experiment 1 found that control decisions were based on one's judgments of learning only in cases of successful retrieval and commission errors. In Experiment 2, judgments of learning were related to restudy decisions only in cases of successful retrieval. For correct answers (Experiments 1 and 2) and commission errors (Experiment 2), individuals were more likely to select the word pair to be restudied when they had assigned it a lower JOL.

This pattern of results may be relevant for both theories of self-regulated learning. When an answer is retrieved, individuals may base monitoring on cues such as the fluency of retrieval (e.g., Schwartz, 1994) to determine the certainty of their answer. For recalled items, lower JOLs may have indicated less certainty in the correctness of the answer. Restudying those items assigned lower JOLs was a strategy whereby individuals may be attempting to reduce the discrepancy between their current and desired learning state (i.e., a strategy in line with the the discrepancy reduction model). In the present study, individuals chose to restudy word pairs for which they had provided an answer on the initial test. Of the items selected for restudy, approximately half were items participants had either answered correctly or answered incorrectly (49% in Experiment 1 and 56% in Experiment 2). By selecting an item for which they have provided an answer at all, participants are selecting an item that is closer to being learned (i.e., items within the region of proximal learning that need more study).

In both experiments, JOLs were not predictive of subsequent study selections for unrecalled items on the initial test. In our study, unrecalled items were less frequently rated with the highest values of JOLs (see Fig. 1). Similarly, correctly recalled items were less frequently rated with the lowest values of JOLs (i.e., zero). When monitoring follows every recall attempt, the lowest judgments are reserved for unrecalled items and the highest for recalled items. However, we note that when a recall attempt immediately precedes monitoring, individuals may use the monitoring scale differently than when no explicit recall attempt occurs (i.e., standard delayed JOLs) or when monitoring does not follow every recall attempt (i.e., feeling of knowing; FOKs). As an illustration, when monitoring follows only unsuccessful recall attempts (standard FOKs), individuals assign the highest values to unrecalled items (Schwartz, Boduroglu, & Tekcan, 2016). Scale differences could lead individuals to adopt different learning strategies. Future studies could explore how control decisions change when monitoring does not follow every recall attempt or when monitoring follows covert retrieval compared to overt retrieval.

Age differences were not present in the relationship between retrieval, monitoring, and control, a finding in line with research that has found the relationship between confidence judgments at the time of test and subsequent control is spared in aging (Hines et al., 2009). We extended this research by demonstrating that JOLs at the time of test are relevant for subsequent control and are used similarly by young and older adults. We note that in Experiment 1 we did find age differences in resolution between monitoring and subsequent restudy selection, as has been previously reported (Dunlosky & Connor, 1997; Souchay & Isingrini, 2004). However, in the mixed effects analyses, age differences in the relationship between judgements of learning and restudy were found only for correctly recalled items. This effect was only marginally significant. Future research could investigate when and why older adults studied correctly recalled items relative to younger adults.

*Future directions and limitations*

In our study, some older adults were more likely to select recalled items. Here, we reference the number of subjects who demonstrated a positive relationship between initial test performance and restudy selection (analyzed in the Supplement) to address the uniformity of strategy adoption (e.g., Dunlosky & Hertzog, 1997). In Experiment 1, nine older adults demonstrated monitoring resolution that was positive. That is, one-third of the older adult participants tested were more likely to select items they had recalled on the initial test. We posit that this strategy adoption arose because the simultaneous presentation of the cues during the restudy selection phase may have increased the chances of evaluating differences in assessed learning among certain items. Further, restudy selections were made after a delay during a separate restudy selection phase, and at the time of selections, participants may have attempted retrieval again. The possibility of attempted retrieval at the time of selection, combined with the ability to make comparisons between all items, may have resulted in different strategic regulation patterns in older adults.

When the task was administered on a trial-by-trial basis during the test, the feedback immediately available for each item would have been emphasized (Koriat et al., 2006). In that case, only two older adults demonstrated monitoring resolution between initial test performance and restudy selection that was positive. Similar to the delayed–JOL effect, where cue–only judgments are less accurate when made immediately (Dunlosky & Nelson, 1994), restudy decisions made immediately may differ than those made after a delay, because of increased reliance on salient cues (i.e., the success of the recall attempt). At minimum, these insights suggest that determining when and why individual differences in strategy adoption depend on the timing and manner of presentation of the control decision is a viable potential future direction.

Differences between Experiments 1 and 2 suggest that a shift in reliance on cues occurred when control was implemented in a sequential manner during the initial testing phase. In contrast to Experiment 1, young and older adults in Experiment 2 were less likely to select items for which they had produced an inaccurate response on the initial test than unanswered items. Second, young adults were less likely to select items for which they had accessed correct partial information as compared to items where no partial information was accessed. In our study, the format of the control decision meaningfully impacted the execution of self-regulated learning, mirroring recent research (Middlebrooks & Castel, 2017). An important limitation of the present results is that it is unknown whether the timing of the restudy decision (immediate versus delayed) or the presentation format (simultaneous versus sequential) were responsible for differences observed between Experiments 1 and 2. Future research should consider factorially manipulating the presentation of the selection phase and timing of the control decision to evaluate its combined effect on self-regulated learning.

Finally, it is important to note some additional limitations. First, the participants in this experiment represented high-functioning older adults with good verbal knowledge. With older adults with lower and larger variation in education levels, it is possible one might see greater effects of age on metacognitive control. Second, the focus of this study was on metacognitive experiences underlying retrieval and their consequences for subsequent SRL. As a part of the design, we did not incorporate a manipulation to explicitly test the efficacy of restudying. While an analysis of final test performance is included in the Supplement, an experiment that directly tests SRL effectiveness, for example, through an honor/dishonor paradigm (Kornell & Metcalfe, 2006) is needed.

## Conclusions

This study considered self-regulated learning within the context of monitoring explicit retrieval. Differences in confidence emerged across the types of responses on the initial test, and the relationship between monitoring and control depended on these types of responses. When the control option occurred during the initial test, commission errors were selected less than items that received no response on the initial test. Whether partial information had been successfully accessed was relevant for restudy selection and indicated divergent strategies between young and older adults. Older adults used access to partial information adaptively, unveiling a novel cue beneficial to older adults' SRL.

## Acknowledgements

The experiments reported were included in the first author's Master's Thesis. Portions of this work were presented at the 57th Annual Meeting of the Psychonomic Society in Boston, MA.

## Appendix A

| Test Items | | Practice Items | |
|---|---|---|---|
| CUE | TARGET | CUE | TARGET |
| Animal | Hope | Cape | Happy |
| Apple | Cheer | Toast | Laugh |
| Bark | Honor | People | Smile |
| Belt | Trust | Plum | Doom |
| Captain | Fright | Report | Suicide |
| Chair | Bliss | Station | Abuse |
| City | Blame | | |
| Copy | Corpse | | |
| Couch | Life | | |
| Doll | Joke | | |
| Drive | Guilt | | |
| Guide | Sick | | |
| Head | Vitality | | |
| Hill | Calm | | |
| Lamp | Sunshine | | |
| Movie | Grief | | |
| Option | Curse | | |
| Planet | Grave | | |
| Ring | Warmth | | |
| School | Flower | | |
| Shade | Award | | |
| Shelf | Hug | | |
| Shirt | Damage | | |
| Shoe | Poverty | | |
| Staff | Filth | | |
| Stand | Loss | | |
| Taste | Love | | |
| Train | Failure | | |
| Tree | Regret | | |
| Walk | Vomit | | |

## Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.04.002.

## References

American Psychological Association (2014). Guidelines for psychological practice with older adults. *American Psychologist, 69*(1), 34–65. https://doi.org/10.1037/a0035063.

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*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.). *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings.* [Technical Report C-1]University of Florida.

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

Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring memory in old age: Impaired, spared, and aware. In J. Dunlosky, & S. K. Tauber (Eds.). *The Oxford handbook of metamemory* (pp. 463–483). New York: Oxford University Press.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 828–835.

Core Team, R. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

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

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 52*(4), 178–186.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.). *Metacognition in educational theory and practice* (pp. 249–276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*(4), 374–380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgements of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*, 545.

Dunlosky, J., Serra, M. J., & Baker, J. M. (2007). Metamemory applied. In F. T. Durso (Vol. Ed.), *Handbook of applied cognition: Vol. 2*, (pp. 137–161). New York: John Wiley & Sons.

Grober, E., Sliwinsk, M., & Korey, S. R. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology, 13*(6), 933–949.

Hanczakowski, M., Zawadzka, K., & Cockcroft-McKay, C. (2014). Feeling of knowing and restudy choices. *Psychonomic Bulletin & Review, 21*(6), 1617–1622. https://doi.org/10.3758/s13423-014-0619-0.

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

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition, 38*(6), 771–784. https://doi.org/10.3758/MC.38.6.771.

Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & Cognition, 42*(1), 126–140. https://doi.org/10.3758/s13421-013-0348-z.

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

Hertzog, C., & Touron, D. R. (2011). Age differences in memory retrieval shift: Governed by feeling-of-knowing? *Psychology and Aging, 26*(3), 647–660.

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. https://doi.org/10.1080/13825585.2015.1028884.

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*(2), 462–475. https://doi.org/10.1037/a0014417.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40*(4), 514–527.

Jacoby, L. L., & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science, 15*(2), 49–53.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469–486. https://doi.org/10.1037/a0017341.

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 923–954.

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

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1095–1105. https://doi.org/10.1037/0278-7393.29.6.1095.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 609–622. https://doi.org/10.1037/0278-7393.32.3.609.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*(5), 493–501.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Dartmouth Publishing Group.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 464–470.

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

Marsh, E. J., Dolan, P. O., Balota, D. A., & Roediger, H. L. (2004). Part-set cuing effects in younger and older adults. *Psychology and Aging, 19*(1), 134–144. https://doi.org/10.1037/0882-7974.19.1.134.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*(3), 349–363. https://doi.org/10.1037/0096-3445.131.3.349.

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1084–1097.

Middlebrooks, C. D., & Castel, A. D. (2017). Self-regulated learning of important information under sequential and simultaneous encoding conditions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(5), 779–792.

Miles, J. R., & Stine-Morrow, E. A. L. (2004). Adult age differences in self-regulated learning from reading sentences. *Psychology and Aging, 19*(4), 626–636. https://doi.org/10.1037/0882-7974.19.4.626.

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition, 29*(Suppl. C), 131–140. https://doi.org/10.1016/j.concog.2014.08.008.

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*(5), 1287–1306. https://doi.org/10.1037/a0036914.

Narens, L., Jameson, K. A., & Lee, V. A. (1994). Subthreshold priming and memory monitoring. In J. Metcalfe, & A. P. Shimamura (Eds.). *Metacognition: Knowing about knowing* (pp. 71–92). Cambridge, MA: MIT Press.

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

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676–686. https://doi.org/10.1037/0278-7393.14.4.676.

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

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bowers (Vol. Ed.), *The psychology of learning and motivation: Advances in research and theory: Vol. 26*, (pp. 125–173). San Diego, CA: Academic Press.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*(1), 53–69. https://doi.org/10.1037/1082-989X.9.1.53.

Nelson, H. E., & Willison, J. (1991). *National adult reading test (NART).* Nfer-Nelson Windsor.

Price, J., & Murray, R. G. (2012). The region of proximal learning heuristic and adult age differences in self-regulated learning. *Psychology and Aging, 27*(4), 1120–1129.

Rich, P. R., Van Loon, M. H., Dunlosky, J., & Zaragoza, M. S. (2017). Belief in corrective feedback for common misconceptions: Implications for knowledge revision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(3), 492–501. https://doi.org/10.1037/xlm0000322.

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

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*(3), 357–375.

Schwartz, B. L., Boduroglu, A., & Tekcan, A.İ. (2016). Methodological concerns: the feeling-of-knowing task affects resolution. *Metacognition and Learning, 11*(3), 305–316. https://doi.org/10.1007/s11409-015-9152-4.

Schwartz, B. L., Pillot, M., & Bacon, E. (2014). Contextual information influences the feeling of knowing in episodic memory. *Consciousness and Cognition, 29*, 96–104. https://doi.org/10.1016/j.concog.2014.08.018.

Sitzman, D. M., Rhodes, M. G., & Tauber, S. K. (2014). Prior knowledge is more predictive of error correction than subjective confidence. *Memory & Cognition, 42*(1), 84–96. https://doi.org/10.3758/s13421-013-0344-3.

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition, 33*(6), 1116–1129.

Souchay, C., & Isingrini, M. (2004). Age-related differences in the relation between monitoring and control of learning. *Experimental Aging Research, 30*(2), 179–193. https://doi.org/10.1080/03610730490274248.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024–1037. https://doi.org/10.1037/0278-7393.25.4.1024.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2011). Context affects feeling of knowing accuracy in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 96–108. https://doi.org/10.1037/a0021612.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2012). An analysis of the determinants of the feeling of knowing. *Consciousness and Cognition, 21*(4), 1681–1694. https://doi.org/10.1016/j.concog.2012.09.005.

Thomas, A. K., Lee, M., & Balota, D. A. (2013). Metacognitive monitoring and dementia: How intrinsic and extrinsic cues influence judgments of learning in people with early-stage Alzheimer's disease. *Neuropsychology, 27*(4), 452–463.

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review, 19*(4), 743–749. https://doi.org/10.3758/s13423-012-0266-2.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x.

Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1073.

Zachary, R. A. (1991). *Shipley institute of living scale.* WPS, Western Psychological Services.