Psychology and Aging (in press)
Accepted 22 November 2020

Should I Keep Studying? Consequences of a Decision to Stop Learning in Young and Older Adults

Aleksandra Krogulska[1,2], Karolina Golik[1], Krystian Barzykowski[1], Jennifer Cox[2], Agata Jakubiak[2], and Elizabeth A. Maylor[2]

[1]Jagiellonian University, Poland

[2]University of Warwick, United Kingdom

Author Note

Aleksandra Krogulska, Institute of Psychology, Jagiellonian University and Department of Psychology, University of Warwick; Karolina Golik and Krystian Barzykowski, Institute of Psychology, Jagiellonian University; Jennifer Cox, Agata Jakubiak, and Elizabeth A. Maylor, Department of Psychology, University of Warwick.

We thank Adrianna Woltmann and Magdalena Kękuś for their assistance with testing older participants in Experiment 1, Saara Khan for her contribution to recruiting and testing participants in Experiment 2, and Katarzyna Zawadzka and Maciej Hanczakowski for their helpful comments on an earlier draft of this manuscript.

This work was funded through a University of Warwick EU Chancellor's Scholarship awarded to Aleksandra Krogulska and partially supported by two National Science Centre (Poland) grants: UMO-2013/09/B/HS6/02642, awarded to Aleksandra Krogulska, and UMO-2015/19/D/HS6/00641, awarded to Krystian Barzykowski. The complete dataset can be found at https://osf.io/4gwjk/

Correspondence concerning this article should be addressed to Aleksandra Krogulska or Elizabeth A. Maylor, Department of Psychology, University of Warwick, Coventry, CV4 7AL. Email: aleksandra.krogulska@warwick.ac.uk, e.a.maylor@warwick.ac.uk.

Abstract

In situations of cognitive overload, the role of a metacognitive decision to stop learning is of utmost importance. We investigated how young and older adults decide to stop learning as a strategy for maximizing memory performance when they face to-be-learned material exceeding their memory capability. People may decide to stop learning for two main reasons: they experience a growing feeling of disfluency as a learning episode progresses and/or they perceive such a decision to be beneficial for future memory performance. In Experiments 1 and 2, participants studied lists of 50 words. The majority of young and older adults stopped learning in conditions where they were allowed to do so. This decision, counterintuitively, decreased the number of recalled words. Crucially, a similar number of young and older adults stopped the presentation of to-be-remembered material, and both age groups suffered comparable consequences in their memory performance. In Experiments 3a and 3b, participants read an experimental scenario and decided whether they would stop learning based on this description alone. People in different age groups predicted their metacognitive decisions similarly. However, participants' forecasted performance did not reflect the negative influence of these decisions. Regardless of their age, people tend to make a suboptimal decision to stop learning, unaware of its negative consequences. Together, our results suggest that young and older adults can exert metamemory control to similar degrees even though their decisions may not be beneficial for memory performance.

Keywords: self-regulated learning, stopping rule, metamemory, cognitive overload, aging

Should I Keep Studying? Consequences of a Decision to Stop Learning in Young and Older Adults

To meet various demands in everyday life, we have to learn fast in a limited time. When we start a new diet, change our job, or explore a new city, the amount of information we need to remember frequently exceeds our memory capacity. The feeling of information overload may become intense, particularly as we get older. Because the proper management of learning activities is an important survival tool (R. A. Bjork, Dunlosky, & Kornell, 2013), people, regardless of age, have developed numerous strategies for dealing with a deluge of information. They tend to prioritize to-be-remembered materials by focusing on essential aspects (Castel et al., 2011; Price, Hertzog, & Dunlosky, 2010) or allocating study time depending on their learning goals (Dunlosky & Ariel, 2011). However, the amount of information that someone is going to encode may determine the choice of learning strategy. Learners must first decide whether to reduce the inflow of new information at some point or try to go through the whole material even if they cannot remember everything. Our research aimed to explore how young and older adults decide to stop studying material before it is presented in its entirety, and what consequences these decisions have for memory retention.

Deciding in any memory task would not be possible without a series of second-order processes called ‘metamemory’, consisting of monitoring and controlling of memory operations as well as applying one’s theories about remembering (Dunlosky & Bjork, 2008; Nelson & Narens, 1990). During each stage of the memory process – encoding, retention and retrieval – three primary metamemory controlling operations can be undertaken: initiation, continuation, and termination of an action (Nelson & Narens, 1990). In a self-regulated learning episode, those decisions boil down to a choice of a moment to start studying, selection of materials, and learning termination. Kornell and Finn (2016) called them ‘large-scale study decisions’ because they determine fundamental aspects of a learning episode and

influence small-scale choices (e.g., which topic to study next, when to stop allocating time to a given item). In the laboratory, researchers curtail people's control over these learning decisions by determining the kind and amount of to-be-remembered materials and study time (Nelson & Narens, 1994). Participants are usually asked to study carefully-chosen materials in their entirety, and they cannot stop learning when they wish. However, carefully controlling a possible source of bias in memory research should not outweigh benefits from investigating the basic, self-directed decisions, including learning termination, that all learners must make outside the laboratory. Mainly, when to-be-remembered materials exceed someone's memory capabilities, from a practical standpoint it is crucial to determine whether people of different ages can make beneficial decisions to stop learning.

Studies exploring learning termination have so far focused on young adults or children (Dufresne & Kobasigawa, 1989; Karpicke, 2009; Kornell & Bjork, 2008; Le Ny, Denhiere, & Le Taillanter, 1972). They show that participants frequently stop learning prematurely. For instance, when students were instructed to learn a list of number-letter associations perfectly, they did not devote enough time and a sufficient number of study-test cycles that would be needed to master all the items (Le Ny et al., 1972). Also, when young participants studied material using flashcards, they tended to drop some flashcards from further restudy even though additional learning could have been beneficial for their memory performance (Kornell & Bjork, 2008).

These observations were bolstered by a recent study by Murayama, Blake, Kerr, and Castel (2016), who examined the consequences of participants' decisions to terminate studying a long list of words. In contrast to earlier studies, where a limited number of to-be-remembered items were presented in a loop, lists were displayed such that each word was presented at most once. In this setup, a decision to stop learning implies that some to-be-learned items will not be presented to a given person at all. Participants were instructed that

the study goal was to recall correctly as many words from the entire list as they could. Across four experiments, more than half of the participants decided to stop the presentation before the end of the list in a condition where they were allowed to do so (in the control condition, participants were presented with all to-be-learned words). Participants who were free to choose the 'stop' option showed worse memory performance – a lower number of free-recalled words – compared to participants in the control condition. Murayama and colleagues also convincingly showed that the impaired memory performance in the stop condition could not be explained by higher task demands created by the active engagement in a decision about when to stop presentation of a word list.[1] When there is more to-be-remembered material than we are actually able to learn, it seems the best strategy is to try to encode as much as possible, even if it causes a feeling of being overwhelmed. These observations accord with the research tradition that stresses the beneficial influence of 'desirable difficulties' experienced during learning. In general, learners frequently choose effortless ways of studying, whereas more strenuous learning strategies enhance long-term memory retention (R. A. Bjork et al., 2013).

Premature decisions to stop studying in young adults probably come from flaws in metamemory abilities – a metamemory control decision maladjusted to the learning goal (Kornell & Bjork, 2008). Should we, then, expect these shortcomings in metamemory control and self-regulated learning to be even more pronounced when we get older? Research on age-related changes in metamemory control during encoding new materials presents an inconclusive picture. There are several studies demonstrating that such control can be

[1] For example, in their Experiment 3, Murayama et al. (2016) compared a standard stop condition with a 'yoked' control condition, where the number of presented words had been matched with another participant's in the stop condition. In this way, participants in the yoked condition could not be distracted by the decision to stop the word presentation, yet were presented with the same number of words as in the stop condition. The authors assumed that if engaging in decision making processes alone actually decreased memory performance, participants in the stop condition would recall a lower number of words in comparison to those in the yoked control condition. Nevertheless, participants recalled a similar number of words in the two conditions.

effectively exercised by older adults and, therefore, serve as a basis for compensation of the decline in their memory performance. For example, older adults can manage their study time during self-paced learning and focus on the most valuable items during value-directed remembering tasks equally well as young adults (Castel et al., 2011; Castel, Murayama, Friedman, McGillivray, & Link, 2013; Hines, Touron, & Hertzog, 2009). Moreover, older adults benefit considerably from task experience. They can update their metamemory judgments effectively and use them strategically to remember important information (McGillivray & Castel, 2011). Hertzog and Hultsch (2000), however, argued that although older adults are capable of effective monitoring of their memory processes, they do not always use that information to regulate and control their learning effectively. Because of the discrepancies emerging from the research on metacognitive control and aging, we formulated two divergent predictions based on the examination of underlying processes operating during the formulation of metamemory judgments. One prediction was based on the role of memory overload by an excess of to-be-remembered material, and the other stressed participants' beliefs about the task (*cognitive overload* vs. *task perception* hypotheses, respectively).

According to the cue-utilization framework, people do not have direct access to memory traces but use various cues (with differing diagnosticity) that may help them to predict future memory performance or assess their learning progress (Benjamin, Bjork, & Schwartz, 1998; Koriat, 1997; Koriat, Bjork, Sheffer, & Bar, 2004). The framework also assumes that metamemory judgments can be derived from two distinctive classes of processes: either a heuristic, nonanalytic one, based on a subjective feeling (experience-based judgments), or deliberate inference from own's beliefs and theories about a given task and memorial consequences of a given learning strategy (theory-based judgments).

Experience-based metamemory judgments depend on online processing of items and are influenced by mnemonic cues such as item familiarity or perceived fluency. Fluency

during encoding and retrieval, defined as easiness of item processing, sometimes is diagnostical for future memory performance (e.g., Schacter & Worling, 1985) but, frequently, it is an unreliable predictor of success in a test (e.g., Benjamin et al., 1998). In the context of the current research, rising difficulty in updating memory with new items may increase encoding disfluency and augment the feeling of being overwhelmed. This, in turn, may prompt people to terminate encoding new items. Experiencing disfluency may be enhanced by deficits in some aspects of long-term and working memory, which may appear when a person gets older (Hasher, Zacks, & May, 1999; Healey, Campbell, & Hasher, 2008). In comparison with young adults, older adults exhibit a decline in episodic memory (Hertzog, Dixon, Hultsch, & MacDonald, 2003), a deficit in associative learning (Old & Naveh-Benjamin, 2008), decreased processing speed (Salthouse, 1996, 2019), and inability to inhibit irrelevant information (Hasher & Zacks, 1988; Hasher et al., 1999). Hess (2014) suggested an increase in the costs of cognitive engagement with age resulting from experiencing cognitive difficulties. Therefore, older adults are much more selective in investing their cognitive resources but once they decide to undertake a given task, they tend to devote a significant amount of cognitive resources to solving it. With sufficient motivation, this pattern of selective but deep investment may attenuate the observed age-related differences. However, in tasks similar to Murayama et al.'s (2016) paradigm, learning continuation may absorb disproportionately more effort in older than young adults. In particular, because of the aforementioned memory deficits, older adults may experience augmented learning disfluency, and based on this mnemonic cue, they may decide to stop studying even earlier than young adults. If, during the study episode, older adults would stop learning more frequently and/or earlier than young adults, observed age-related differences in memory performance may widen.

In contrast, theory-based metamemory judgments reflect deliberate use of beliefs and memories to draw conclusions about one's own competence or beneficial strategies in solving a given task (Koriat et al., 2004). For example, an anticipated test format influences the way in which learners regulate their study behavior (Finley & Benjamin, 2012). Previous studies have shown that young and older adults have similar assumptions about the usefulness of different learning strategies (Castel et al., 2011). Moreover, a recent study on metamemory beliefs showed only a small effect of age on the perception of learning strategies in a group of middle-aged and older adults from different cohorts (Hertzog, Small, McFall, & Dixon, 2019). Hence, it is also possible that young and older adults would perceive a decision to terminate learning similarly and would use the 'stop' option in the same way as would young adults.

To give an overview of our research: Experiments 1 and 2 compared young and older adults' decisions to stop learning, and their consequences for memory performance. Such decisions have not been explored so far in the older adult population, but they may add valuable input to our understanding of developmental aspects of self-regulated learning. Furthermore, the comparison of young and older adults may shed light on possible stopping mechanisms: whether they are based mostly on experiencing mnemonic cues (e.g., learning fluency) or participants' beliefs regarding consequences of stopping. Participants' perception of the task was also directly investigated in Experiments 3a and 3b, where they predicted whether they would stop learning based on the task description alone.

Experiment 1

In Experiment 1, we adapted Murayama et al.'s (2016) paradigm to validate their observations in different age groups. Young and older participants studied lists of 50 words. In the control condition, participants studied entire lists, whereas in the stop condition participants could stop the presentation of each list before the end, their task being to recall as

many words from the entire list as they could. Based on the cognitive overload hypothesis, we expected that more older participants should decide to stop the material, and also earlier, than young participants, and suffer stronger reductions in recall. In contrast, the task perception hypothesis predicts that comparable numbers of young and older participants should decide to stop learning, and at similar points, with equivalent consequences for recall.

Participants completed three study-test cycles, which allowed us to investigate how the consequences of learning termination change across study-test cycles. On the one hand, participants' experience and updated knowledge about a task may help them in making more profitable decisions when learning subsequent word lists. In consequence, they may avoid the 'stop' option or use it adaptively, that is, without hampering memory performance. On the other hand, participants – especially older adults – may become more tired and overwhelmed as they progress with the task. Such conditions may prompt participants to stop learning earlier as the study-test cycles progress.

Method

Participants. Sixty-four students (57 female, age: 19-38, *M* = 22.23, *SD* = 3.55) of the Jagiellonian University in Cracow and 64 older adults (59 female, age: 60-85, *M* = 66.41, *SD* = 4.68) took part in the study. In cognitive aging research, one common issue with an extreme-groups design is using college students involved in various mental activities as a reference group, possibly resulting in biased estimates of differences between young and older adults. Therefore, to find a suitable match for cognitively active students, most of the older adults (above 80%) were recruited from the University of the Third Age. Murayama et al.'s (2016) main effect of experimental group (control vs. stop) on the number of correctly recalled words had an approximate effect size of $\eta_p^2 = .061$.[2] A sample size equal to 42

[2] Exact numbers of η_p^2 for Experiments 1, 2, and 4 of Murayama et al. (2016) were .063, .069, and .050, respectively. These values were calculated based on datasets kindly provided by the first author.

participants would be sufficient to obtain a comparable effect with a statistical power of .95 and α-level at .05. However, we decided to test 64 participants per age group to be able to detect possibly smaller effect sizes in a hitherto unexplored older adult age group. Thus, the control and stop groups each consisted of 32 young and 32 older participants, who were assigned to the groups randomly.

Inclusion criteria for older participants were: minimal age of 60 years, native fluency in Polish, finished at least secondary-school, no history of dementia-related problems, and no use of medication affecting cognition. Moreover, after the experimental session, all older adults filled in the Addenbrooke's Cognitive Examination Revised (ACE-R) test battery. All older participants scored within the norms or above; this excludes with a substantial probability the occurrence of dementia symptoms in our sample.

All participants were treated in accordance with the British Psychological Society Code of Human Research Ethics and the Polish National Science Centre guidelines for research ethics. Participants received no incentives.

Materials. In total, 150 Polish nouns of 4-6 letters in length were chosen from the SUBLEX database (Mandera, Keuleers, Wodniecka, & Brysbaert, 2015). All words had average lexical frequency (2.19-5.78, M = 4.07)[3] and were emotionally neutral. They were split into three lists consisting of 50 words each, with assignment of words to lists fixed. List order was counterbalanced across participants.

Procedure. Testing took place in sound-proof cubicles in a university laboratory. Participants were informed that they would be presented with three lists of words, but they were not informed about the number of words in each list. They were instructed that their goal was to recall as many words as possible from each list after its presentation. During the

[3] The measure of word frequency is presented on the Zipf scale, which is independent of the corpus size. For a majority of words, this measure takes a value between 1 and 7 (Mandera et al., 2015).

study phase, the words were displayed (font: Arial; size: 52) one-by-one on a computer screen in random order for 2 s per word, with a 500-ms interstimulus interval. Immediately after the study phase, participants completed a short numerical filler task in which they counted down aloud from 495 by 7s for 15 s. Finally, in the test phase, they were instructed to write on a sheet of paper as many words as they could remember from the list just presented. The free-recall test took two minutes. Following the test phase, the procedure progressed to the next study-test cycle. When all cycles had been completed, participants received debriefing.

For the control group, all three 50-word lists were displayed in their entirety, whereas participants in the stop group had the option to stop each list of words before it had been fully shown. Thus, during the study phases, a checkbox labelled 'End list' was displayed on the computer screen, and participants were able to activate it by pressing the Enter key (which was clearly labelled 'End list') on the keyboard. They were instructed that this was not a mandatory option, and they should use it only when they believed that this was an optimal strategy to remember as many words as possible. When participants pressed the Enter key, the computer program skipped the remainder of the list, and proceeded straight to the counting task and, then, to the memory test.

Results

'Stop' Decisions. Overall, more than half of the participants decided to terminate their study phases when they were allowed to do so – on average, 58.3% of young and 67.7% of older adults (see Table 1). These age groups did not significantly differ in their propensity to stop: similar number of young and older adults decided to stop List 1, $\chi^2(1) = 0.07$, $p = .798$; List 2, $\chi^2(1) = 0.61$, $p = .434$; and List 3, $\chi^2(1) = 1.70$, $p = .193$. Willingness to terminate learning was also stable across study-test cycles. Cochran's Q test did not indicate

any differences among Lists 1, 2 and 3 in the number of participants who decided to stop, $\chi^2(2) = 0.50$, $p = .779$ (young); and $\chi^2(2) = 2.80$, $p = .247$ (older).

'Stop' Serial Positions. The mean serial positions at which participants stopped learning in the stop group are presented in Table 2. If participants chose not to stop, this was coded as serial position 50. These positions were analyzed as a function of participants' age (young vs. older) and list (1, 2, 3), using a two-factor mixed ANOVA. This did not yield a significant main effect of age, $F < 1$, $\eta_p^2 = .003$, nor a main effect of list, $F < 1$, $\eta_p^2 = .015$. These effects, however, were qualified by a significant interaction between age and list, $F(2, 124) = 5.47$, $MSE = 92.86$, $p = .005$, $\eta_p^2 = .081$.

To disentangle this interaction, we examined the effect of list separately for young and older adults. A one-way repeated measures ANOVA did not indicate any difference in the position at which young adults stopped learning across lists, $F(2, 62) = 1.17$, $MSE = 114.69$, $p = .318$, $\eta_p^2 = .036$. However, the effect of list proved to be significant in the case of older adults, $F(2, 62) = 6.48$, $MSE = 71.04$, $p = .003$, $\eta_p^2 = .173$. Pairwise comparisons (with the Bonferroni adjustment) showed that, as the study progressed across lists, older adults were stopping learning at earlier positions, with a significant difference between Lists 1 and 3 of 7.22, $SE = 2.40$, $p = .015$, and a numerical trend in the same direction between Lists 1 and 2 of 5.63, $SE = 2.60$, $p = .114$.

Correctly Recalled Words. Figure 1 illustrates the mean numbers of correctly recalled words by young and older adults.[4] Participants' answers were treated as correct if

[4] We followed Murayama et al. (2016) in focusing here on quantity of recall. However, we also analyzed the recall data in terms of quality of recall, that is, response accuracy (the number of correctly recalled words divided by the total number of words recalled). We suspected that, although participants were explicitly instructed that their goal was to recall as many words as possible, they may, nonetheless, be partially driven by a willingness to provide only correct responses in the memory test. Simply, they might have treated 'stop' decisions as a helpful strategy to maintain high accuracy of their responses. In that case, participants in the stop group would be more accurate than in the control group. To summarize, in both Experiments 1 and 2, older adults were significantly less accurate in their recall than were young adults but there were no significant differences between the stop and control conditions (see Supplementary Materials for full details).

they matched words on corresponding lists; instances of small spelling errors (e.g., *lizzard* instead of *lizard*) or pluralization (e.g., *bells* instead of *bell*) were accepted. Like Murayama et al. (2016), to investigate the consequences of learning termination, we compared young and older adults' memory performance in the stop and control groups using an ANOVA. Next, we examined correlations between the number of correctly recalled words and the stop position. These two methods together allow us to assess whether learning termination impairs memory performance, and whether the degree of this impairment is related to the position at which participants stopped.

A 2 (age) × 2 (group: stop vs. control) × 3 (list) mixed ANOVA on the number of correctly recalled words yielded a significant main effect of group, $F(1, 124) = 5.63$, $MSE = 48.83$, $p = .019$, $\eta_p^2 = .043$, indicating that participants in the control group ($M = 9.62$, $SD = 4.66$) correctly recalled more words than those in the stop group ($M = 7.92$, $SD = 5.13$). A significant main effect of age, $F(1, 124) = 61.86$, $MSE = 48.83$, $p < .001$, $\eta_p^2 = .333$, indicated that young adults ($M = 11.57$, $SD = 5.15$) correctly recalled more words than older adults ($M = 5.96$, $SD = 2.65$). Crucially, the interaction between age and group was not significant, $F < 1$, $\eta_p^2 < .001$. Thus, the consequences of having the 'stop' option at participants' disposal were comparable for young and older adults' memory performance. The interaction between list and group, as well as the three-way interaction, were non-significant: $F(2, 248) = 1.59$, $MSE = 8.67$, $p = .206$, $\eta_p^2 = .013$; and $F(2, 248) = 1.52$, $MSE = 8.67$, $p = .220$, $\eta_p^2 = .012$; respectively.[5]

[5] We repeated the same analysis for all recalled words to give an overview of total recall (see Table A1 in the appendix for means). The results mimic those obtained on the numbers of correctly recalled words. Here, we present only results related to the consequences of learning termination. A 2 (age) × 2 (group: stop vs. control) × 3 (list) mixed ANOVA on the total number of words yielded a significant main effect of group, $F(1, 124) = 9.43$, $MSE = 48.97$, $p = .003$, $\eta_p^2 = .071$. Participants in the control group ($M = 11.00$, $SD = 4.68$) recalled more words than those in the stop group ($M = 8.81$, $SD = 5.08$). There was a significant main effect of age, $F(1, 124) = 60.09$, $MSE = 48.97$, $p < .001$, $\eta_p^2 = .326$. Young adults ($M = 12.67$, $SD = 5.20$) recalled more words than older adults ($M = 7.14$, $SD = 2.75$). The interaction between age and group, the interaction between list and group, as

There was a significant main effect of list, $F(2, 248) = 3.69$, $MSE = 8.67$, $p = .026$, $\eta_p^2 = .029$, but this was qualified by a significant interaction with age, $F(2, 248) = 3.06$, $MSE = 8.67$, $p = .049$, $\eta_p^2 = .024$. To explore this further, we conducted two-way ANOVAs, separately for young and older adults. For young adults, the increase in recall across lists was significant, $F(2, 124) = 4.76$, $MSE = 12.10$, $p = .010$, $\eta_p^2 = .071$, with no interaction with group, $F(2, 124) = 2.23$, $MSE = 12.10$, $p = .112$, $\eta_p^2 = .035$. In contrast, for older adults, there was no effect of list, $F < 1$, $\eta_p^2 = .003$, and no interaction with group, $F < 1$, $\eta_p^2 < .001$.

To reinforce the main finding that terminating learning impairs later memory performance, correlations between serial positions at which participants stopped learning and numbers of correctly recalled words in the stop group were performed. Indeed, positive and significant correlations were observed for young adults – List 1, $r(32) = .73$, $p < .001$; List 2, $r(32) = .63$, $p < .001$; List 3, $r(32) = .66$, $p = .001$; and for older adults – List 1, $r(32) = .59$; $p < .001$; List 2, $r(32) = .49$, $p = .005$; List 3, $r(32) = .61$, $p < .001$. These indicate that participants who stopped earlier recalled fewer words.

Discussion

The results of Experiment 1 mainly supported the task perception hypothesis and showed that a similar majority of young and older adults terminated learning, and did so at comparable positions, with equivalently detrimental effects on memory performance. To the best of our knowledge, this experiment represents the first attempt to investigate a metamemory decision to terminate learning with the participation of older adults. We also observed some changes in participants' study behavior over the course of completing the three study-test cycles: older adults stopped learning at earlier positions whereas young adults

---

well as the three-way interaction, were non-significant: $F < 1$, $\eta_p^2 < .001$; $F(2, 248) = 1.43$, $MSE = 9.47$, $p = .241$, $\eta_p^2 = .011$; and $F(2, 248) = 1.50$, $MSE = 9.47$, $p = .225$, $\eta_p^2 = .012$, respectively.

performed better with each subsequent memory test. The fact that older adults stopped learning earlier with each list seems to provide some support to the cognitive overload hypothesis, too. Note that this potential contribution of experience-based monitoring was observable because of the experimental design with multiple word-lists. Thus, it is possible that for single-list experiments, task perception may play a more important role. Experiment 2 was designed to provide some additional insight into the underlying mechanisms of learning termination as well as to test how experience with the task may shape learning strategies.

Experiment 2

Our main aim in Experiment 2 was not only to replicate the results of Experiment 1 but also to extend them by employing a different design. In Experiment 2, the degree of participants' control over to-be-learned materials was manipulated via the use of a within-subject design. Participants were asked to study four lists of 50 words. They could stop learning in two out of four word-lists. One experimental group started with the two 'stop' lists (the stop-first group), the other group with the two control lists (the stop-second group). With this arrangement, we were able to additionally examine the influence of participants' knowledge and experience of the task.

Before the experiment commenced, participants did not receive information about the length of the lists. Nevertheless, participants in the stop-second group had a chance to practice learning two entire study-test cycles before they were allowed to make their first 'stop' decision. As a result, they could learn the length of the word-lists and presentation time from their own experience. In the stop-first group, participants were unaware of these features of the study materials. While knowledge about learning conditions can potentially affect participants' metacognitive decisions, this influence can be more pronounced for older adults. They tend to start choosing more beneficial learning strategies when they receive

additional information about the task (Dunlosky & Hertzog, 2001; Murphy et al., 1987). Moreover, previous experience with to-be-learned materials and free-recall tests can improve memory and metamemory performance in young and older adults (E. L. Bjork, de Winstanley, & Storm, 2007; Kuhlmann & Undorf, 2018; Price, Hertzog, & Dunlosky, 2008). Therefore, we expected that young and older adults in the stop-second group would benefit from additional experience and knowledge of the task. Hence, they should avoid the 'stop' option more frequently than participants in the stop-first group. Furthermore, once they decided to stop, they would do that at a later position, consequently suffering less harmful consequences for their memory performance.

Method

Participants. Sixty-four undergraduate students (48 female, age: 17-29, M = 20.22, SD = 2.00) of the University of Warwick and 64 older adults (37 female, age: 60-90, M = 73.50, SD = 7.00) participated in the study. Older adults were independently living in the community and were recruited via the University of the Third Age, senior citizens clubs, churches, etc. The study was approved by the University of Warwick Psychology Ethics Committee. Half of the young adults in each group participated in exchange for course credit. Older adults tested at the university received £10 as a contribution to their travel expenses. Thirty-two young and 32 older participants were randomly assigned to the stop-first group; the remaining participants were assigned to the stop-second group.

All participants were either native English speakers or bilingual with a fluent level of English (one older adult and three young adults). The inclusion criteria were similar to those in Experiment 1, except that we asked participants to provide us with the total number of years of their formal education, which did not differ across age groups (young: M = 15.00, SD = 1.52; older: M = 15.55, SD = 3.40), $t(126) = -1.18$, $p = .242$, $d = 0.21$.

To assess cognitive functioning, participants completed the Digit Symbol Substitution task (Wechsler, 1981) as a measure of processing speed, and the Mill Hill Vocabulary test (Raven, Raven & Court, 1998) as a measure of crystallized intelligence. As expected, young adults (M = 62.34, SD = 10.94) performed significantly better in the processing speed test than did older adults (M = 45.52, SD = 9.96), $t(126) = 9.10$, $p < .001$, $d = 1.61$. The pattern was reversed for vocabulary where young adults (M = 17.17, SD = 3.16) received lower scores than did older adults (M = 23.77, SD = 4.42), $t(126) = -9.71$, $p < .001$, $d = 1.71$.

Materials. In total, 200 English nouns were chosen from the English Lexicon Project website (Balota et al., 2007). The hyperspace analogue to language (HAL) frequency ranged from 8.26 to 10.26 (M = 9.26), corresponding to the frequency reported in Murayama et al.'s (2016) paper. As in Experiment 1, we chose 4-6-letter emotionally neutral words of medium lexical frequency. Words were then split into four lists consisting of 50 words each. The assignment of words to lists was fixed. Two lists were assigned to the stop condition, and the two remaining to the control condition. Assignment of a list to the condition as well as the list order within conditions was fully counterbalanced across participants.

Procedure. Participants were tested individually in quiet rooms, either in their own homes or at the university. The experimental procedure was similar to that of Experiment 1, with a few exceptions. Participants completed four study-test cycles instead of three, and control over to-be-learned materials was manipulated within-subjects rather than between-subjects: Participants in the stop-first group had the opportunity to stop learning in the first two to-be-learned lists (two remaining lists were displayed in their entirety), whereas those in the stop-second group could terminate learning in the second two lists. When all cycles had been completed, participants performed the speed and vocabulary tests.

Results

'Stop' Decisions. In both groups, about 70% of young and older participants decided to terminate learning (see Table 1). As in Experiment 1, a comparable number of young and older adults decided to stop learning in the stop-first group: List 1, $\chi^2(1) = 0.29$, $p = .590$; List 2, $\chi^2(1) = 0.08$, $p = .777$. This was also the case for the stop-second group: List 1, $\chi^2(1) = 1.33$, $p = .248$; List 2, $\chi^2(1) = 1.24$, $p = .266$. Moreover, the stop-first and stop-second groups did not differ in their rates of stopping: young List 1, $\chi^2(1) = 0.08$, $p = .784$; young List 2, $\chi^2(1) = 0.67$, $p = .412$; older List 1, $\chi^2(1) = 2.00$, $p = .157$; older List 2, $\chi^2(1) = 0.33$, $p = .564$. Finally, an exact McNemar's test was performed to check whether the tendency to stop was stable across the two study-test cycles. In both stop-first and stop-second groups, this test determined that there were no statistically significant differences between Lists 1 and 2 in the number of young adults who decided to stop, *p*s = 1. Similarly, a comparable number of older adults stopped learning Lists 1 and 2 in the stop-first group, $p = .50$; and in the stop-second group, $p = 1$.

'Stop' Serial Positions. The mean serial positions at which young and older participants stopped learning are presented in Table 2. These serial positions were analyzed in the stop condition as a function of participants' age (young vs. older), list (1 vs. 2) and experimental group (stop-first vs. stop-second) in a three-factor mixed ANOVA. Unlike Experiment 1, this yielded a significant main effect of age, $F(1, 124) = 5.83$, $MSE = 431.92$, $p = .017$, $\eta_p^2 = .045$. This time, older adults ($M = 26.61$, $SD = 15.42$) stopped learning at earlier positions than did young adults ($M = 32.88$, $SD = 13.88$). A significant main effect of list was also observed, $F(1, 124) = 8.83$, $MSE = 73.26$, $p = .004$, $\eta_p^2 = .066$. This was qualified by a significant interaction between group and list, $F(1, 124) = 4.77$, $MSE = 73.26$, $p = .031$, $\eta_p^2 = .037$, which arose because only in the stop-first group did participants stop earlier in List 2 ($M = 28.03$, $SD = 16.52$) than in List 1 ($M = 33.55$, $SD = 15.21$). The main effect of group was not significant, $F < 1$, $\eta_p^2 = .005$, as, contrary to our expectations, participants stopped

learning at a similar position in the stop-first (*M* = 30.79, *SD* = 14.80) and the stop-second groups (*M* = 28.70, *SD* = 15.14). Also, interactions between list and age, group and age, and the three-way interaction were not significant, *F*s < 1.

Correctly Recalled Words. Figure 1 illustrates the mean numbers of correctly recalled words by young and older adults. A four-factor mixed ANOVA on recall included participants' age (young vs. older), condition (stop vs. control), list (1 vs. 2), and experimental group (stop-first vs. stop-second). The main effect of condition was significant, $F(1, 124) = 30.88$, $MSE = 14.69$, $p < .001$, $\eta_p^2 = .199$, confirming the expected decrease in recalled words in the stop condition (*M* = 7.31, *SD* = 5.53), compared to the control condition (*M* = 9.19, *SD* = 5.57). A significant main effect of age, $F(1, 124) = 42.76$, $MSE = 81.78$, $p < .001$, $\eta_p^2 = .256$, was also observed. Young adults correctly recalled more words (*M* = 10.86, *SD* = 5.83) than did older adults (*M* = 5.64, *SD* = 2.61), while the interaction between age and condition was not significant, $F(1, 124) = 1.13$, $MSE = 14.69$, $p = .291$, $\eta_p^2 = .009$. Thus, termination of learning impaired memory performance equally in the two age groups, replicating Experiment 1. The main effects of group and list were not significant: $F(1, 124) = 1.62$, $MSE = 81.78$, $p = .206$, $\eta_p^2 = .013$, and $F < 1$, $\eta_p^2 = .006$, respectively.[6]

Almost no interactions yielded statistical significance. Crucially, the interaction between condition and group was not significant, $F < 1$, $\eta_p^2 < .001$. Combined with the absence of an overall effect of group, this indicates that, contrary to our expectation, stopping the list early had the same consequences for memory performance in the stop-first and stop-

[6] We replicated the primary effects for the total numbers of recalled words (see Table A1 in the appendix for means). A four-factor mixed ANOVA showed a main effect of condition, $F(1, 124) = 44.06$, $MSE = 15.48$, $p < .001$, $\eta_p^2 = .262$. Participants in the control condition (*M* = 10.63, *SD* = 5.59) recalled more words than those in the stop condition (*M* = 8.32, *SD* = 5.55). There was a significant main effect of age, $F(1, 124) = 39.91$, $MSE = 83.10$, $p < .001$, $\eta_p^2 = .243$. Young adults (*M* = 12.02, *SD* = 5.77) recalled more words than older adults (*M* = 6.93, *SD* = 2.85), with no interaction between age and condition, $F < 1$, $\eta_p^2 = .002$. The main effect of group was also not significant, $F(1, 124) = 1.50$, $MSE = 83.10$, $p = .222$, $\eta_p^2 = .012$.

second groups. Only the three-way interaction between list, condition and group was significant, $F(1, 124) = 4.55$, $MSE = 8.65$, $p = .035$, $\eta_p^2 = .035$. This was attributable to a general 'time-on-task' trend such that recall dropped overall from 8.44 for the first of the four lists, to 8.10 for the second, and to 7.84 for the third, but then rose to 8.62 for the last.[7]

All correlations between the serial position at which the participants decided to stop learning and the number of correctly recalled words proved to be significant in the stop-first group (young adults: List 1, $r(32) = .55$, $p = .001$; List 2, $r(32) = .67$, $p < .001$; older adults: List 1, $r(32) = .46$, $p = .008$; List 2: $r(32) = .44$, $p = .011$), as well as in the stop-second group (young adults: List 1, $r(32) = .43$, $p = .014$; List 2, $r(32) = .48$, $p = .005$; older adults: List 1, $r(32) = .44$, $p = .011$; List 2, $r(32) = .66$, $p < .001$).[8]

Discussion

The results of Experiment 2 successfully replicated those of Experiment 1. The majority of young and older adults terminated presentation of to-be-remembered material despite the fact that the best strategy for maximizing memory performance is to try to encode as much as possible and not avoid learning new materials. Again, the pattern of results mostly supports the task perception rather than the cognitive overload hypothesis. Similar numbers of young and older adults decided to stop and they suffered comparable consequences of this

[7] The overall 'time-on-task' effect indicated that the number of correctly recalled words decreased steadily over the first three lists. This result may be explained by proactive interference (Bjork, 2003). However, the unexpected increase in recall performance on the fourth list is more difficult to explain. It may be a result of gaining some experience with the task (Storm, Hickman, & Bjork, 2016) or benefiting from practicing retrieval in the previous memory tests (*forward testing effect*, Wahlheim, 2015).

[8] To check whether other variables are related to the 'stop' positions, we used demographic information and secondary findings collected in Experiment 2. Separately for each age and experimental group, we conducted correlations between stopping positions on List 1 and List 2 and: age, years of education, processing speed, vocabulary, and correct recall from the control lists. In the stop-second group, we observed significant (or close to significant) correlations between the serial 'stop' position and age (young adults: List 1, $r(32) = -.33$, $p = .062$; List 2, $r(32) = -.42$, $p = .017$; older adults: List 1, $r(32) = -.66$, $p < .001$; List 2: $r(32) = -.46$, $p = .009$), correct recall from the control List 1 (older adults: List 1, $r(32) = .37$, $p = .039$; List 2: $r(32) = .41$, $p = .019$) and correct recall from the control List 2 (young adults: List 1, $r(32) = .41$, $p = .019$; List 2, $r(32) = .55$, $p = .001$; older adults: List 1, $r(32) = .40$, $p = .024$) . In the stop-first group, only the correlation between the serial 'stop' position on List 1 and correct recall on the control List 1 for young adults was close to significance, $r(32) = .32$, $p = .077$. Thus, it seems that after participants gain some experience with the task, their age and memory abilities may be good predictors of the 'stop' position. However, this issue needs a more targeted investigation.

decision in memory tests. Thus, regardless of age, people may assume that restricting the inflow of new information at some point is a beneficial learning strategy and they may not be aware of the negative consequences of doing so. Also, a similar number of young and older adults stopped learning in the stop-first and stop-second groups. Thus, it seems that previous experience and knowledge of the length of the lists had little influence on the 'stop' decisions.

Nonetheless, the role of experience-based metamemory judgments in making the 'stop' decisions should not be ignored. Although a similar number of young and older adults decided to stop in both Experiments 1 and 2, older adults stopped learning earlier than did young adults in Experiment 2. In Experiment 1, older adults also tended to stop earlier in the second and third study-test cycles in comparison with the first study-test cycle. Therefore, it appears that a feeling of being overloaded and/or fatigued may influence learning termination to some degree, at least in the case of older adults.

Experiments 3a and 3b

Although the results of Experiments 1 and 2 mostly supported the task perception hypothesis, we did find some evidence that a feeling of being overloaded plays some role in older adults' decisions to stop. Experiments 3a and 3b were conducted to provide direct evidence that task perception determines termination of learning in different age groups. These studies investigated participants' assumptions and beliefs about the decision to stop learning in a situation when the possibility of experiencing cognitive overload and exercising control over learning had been eliminated.

In Experiments 3a and 3b, young and older participants read a description of a psychological study like our Experiment 1. Their task was to imagine that they were taking part in the described study. Based on each task description alone, participants predicted their memory performance and, when appropriate, decided whether (and where) they would stop

learning (see Murayama et al., 2016, Experiment 5, for a similar manipulation). In Experiment 3b, we tested a middle-aged group as well to check whether beliefs about learning termination fluctuate or are stable across adulthood. If task perception plays the crucial role in learning termination, in Experiments 3a and 3b, comparable numbers of young, middle-aged, and older adults should predict that they would decide to terminate learning. Furthermore, participants' willingness to terminate and the anticipated 'stop' positions should reflect the results obtained in Experiment 1.

Method

Participants. In Experiment 3a, 64 students (54 female, age: 20-31, *M* = 22.98, *SD* = 0.30) of the Jagiellonian University in Cracow and 64 older adults (48 female, age: 60-87, *M* = 72.03, *SD* = 0.81) participated in the study. Most of the older adults were recruited at the University of the Third Age in Cracow. Inclusion criteria and ethical standards were identical to those in Experiment 1. Thirty-two young and 32 older participants were randomly assigned to the control group, and the remaining half of the participants were assigned to the stop group. Two older participants (one in each group) who responded affirmatively to a question about dementia-related problems were excluded from further analysis.

In Experiment 3b, 110 participants aged 18-84 took part in the study. For further analysis, we used only the results from participants who fully completed an online or paper-based version of the task and fell into the desired age ranges (i.e., young: 18-30, middle-aged: 40-55, older: 65-80). Consequently, we analyzed data from 90 participants: 27 young (14 female, *M* = 21.74, *SD* = 2.65), 29 middle-aged (20 female, *M* = 47.07, *SD* = 4.96), and 34 older adults (18 female, *M* = 70.26, *SD* = 4.13). The other inclusion criteria for participants were: not university students, native or high proficiency in English, no history of dementia-related problems, and no need for assistance in everyday life. The study was approved by the University of Warwick Psychology Ethics Committee.

Materials. Descriptions of two experimental scenarios corresponding to the two groups in Experiment 1 were prepared in Polish (Experiment 3a) and in English (Experiment 3b). Each scenario was depicted as a single-list experiment (50 words), to simplify the participants' decision process. Participants were informed that the goal of the described experiment was to recall correctly as many words as possible. In the control group, the participants received a scenario in which participants needed to learn a list of 50 unrelated words. This scenario contained information that each word would be displayed on the computer screen individually, for 2 s each. They were also informed that when the presentation of all the words was over, the procedure would progress to a 15-s task of counting down aloud from 495 by 7s. Finally, they were informed that they would have two minutes to write down as many words as possible from the list. The stop scenario contained the same details, but with the additional information that participants were allowed to stop presentation of the list before it had been presented in its entirety and that they could use this option if they thought that this was a beneficial strategy to maximize their memory performance.

In Experiment 3b, participants received both scenarios and were randomly assigned to the order in which they completed the task. Thus, 46 participants started with the stop scenario (13 young, 15 middle-aged, 16 older) and the remaining 44 started with the control scenario (14 young, 14 middle-aged, 18 older). Moreover, descriptions contained three examples of list items (e.g., *survey*, *robot*, *clay*).

Procedure. In Experiment 3a, participants read the description of a hypothetical memory experiment and were instructed to imagine that they were taking the test. In the control group, they were asked to predict how many words they would recall out of a 50-word list. In the stop group, they were first asked whether they would decide to stop the presentation of a word list in order to maximize the number of correctly recalled words in a

later memory test. If they responded affirmatively, they were asked to indicate after which of the 50 words they would stop the list. Finally, they were asked to predict how many words they would recall.

In Experiment 3b, participants completed either a paper-based (N = 15, in the presence of an experimenter) or online version of the task (N = 75, without an experimenter). This time, each participant received both hypothetical scenarios.[9] After answering questions about these scenarios, participants were asked to produce a written answer to one of the following questions: 'In the Stop scenario, if you chose to stop, why?' or 'If you chose not to stop, why not?', depending on whether they predicted that they would stop the list. Participants then completed a short questionnaire regarding their beliefs about the termination of learning .[10] They needed to think back to the stop scenario and tick a box indicating whether they agreed or disagreed with each of the following statements: 'Ending the list early is a good strategy to help recall as many words as possible', 'Ending the list early will prevent overloading of memory', 'Ending the list early will help ensure that all the responses are correct', and 'Ending the list early will reduce the effort in learning the words'.

Results

Numbers and percentages of participants who predicted they would decide to stop learning, and their predicted stopping positions are presented in Table 3.

Predicted 'Stop' Decisions. In Experiment 3a, more than half of the participants predicted that they would have terminated learning. Numerically more older adults than young ones declared that they would have stopped, but this difference did not reach statistical

[9] Note that in both experiments, participants ascertained the number of to-be-learned words (50) from the experimental instruction and might have made a hypothetical 'stop' decision based on this knowledge of the length of the list. In contrast, in Experiments 1 and 2, the instruction did not contain information about the number of to-be-learned words. However, in the stop-second group in Experiment 2, participants not only could see how long the list was but they also gained some experience in learning 50-word lists. The results of Experiment 2 showed that additional knowledge about the length of the list neither influenced the participants' decision nor changed the position at which they decided to stop learning.

[10] Detailed description and analysis of participants' answers may be found in the Supplementary Materials.

significance, $\chi^2(1) = 3.77$, $p = .052$. Moreover, there was no significant difference between the number of participants who, in fact, stopped learning in Experiment 1 (on List 1) and participants who predicted that they would have decided to stop on the basis of the task description alone: $\chi^2(1) = 1.00$, $p = .316$, and $\chi^2(1) = 0.51$, $p = .476$, for young and older adults, respectively.

In Experiment 3b, about half of the participants predicted that they would terminate the learning. A similar number of young, middle-aged, and older adults predicted they would stop during learning, $\chi^2(2) = 2.35$, $p = .309$. Moreover, similar numbers of young ($\chi^2(1) = 1.31$, $p = .253$) and older adults ($\chi^2(1) = 0.003$, $p = .760$) stopped learning in Experiment 1 (on List 1) and predicted that they would have done so in Experiment 3b.

To explore further the numerical young-older differences evident in Table 3, we pooled the data across experiments for greater power. With combined numbers, significantly fewer young (27/59; 45.8%) than older adults (43/65; 66.2%) predicted they would terminate learning, $\chi2(1) = 5.23$, $p = .022$.

Predicted 'Stop' Serial Positions. If participants provided a range instead of a specific point on the list (one older adult) the mid-value was used. In Experiment 3a, mean serial position at which participants forecasted to stop was comparable for young and older adults, $t(61) = 1.19$, $p = .239$, $d = .299$. We then tested whether the predicted 'stop' positions were comparable to those observed in Experiment 1 (List 1). A 2 (age) × 2 (Experiment: 1 vs. 3a) ANOVA on the 'stop' position did not yield a significant main effect of experiment, $F < 1$, $\eta_p^2 = .002$, nor a main effect of age, $F < 1$, $\eta_p^2 < .001$, and no interaction, $F(1, 123) = 2.43$, $MSE = 276.33$, $p = .122$, $\eta_p^2 = .019$.

In Experiment 3b, a one-way ANOVA indicated that the mean serial position at which the participants forecasted to stop learning was comparable among young, middle-aged, and older adults, $F(2, 87) = 1.30$, $MSE = 238.02$, $p = .277$, $\eta_p^2 = .029$. Young and older

adults' predictions were then compared with the actual stop positions observed in Experiment 1 (List 1). A 2 (age: young vs. older) × 2 (Experiment: 1 vs. 3b) ANOVA on the 'stop' positions did not yield a significant main effect of experiment, $F(1, 121) = 1.39$, $MSE = 268.08$, $p = .240$, $\eta_p^2 = .011$, nor a main effect of age, $F < 1$, $\eta_p^2 < .001$, and no interaction, $F(1, 121) = 2.66$, $MSE = 268.08$, $p = .106$, $\eta_p^2 = .022$.

Predicted Memory Performance. (See Figure 2 for means.) In Experiment 3a, a 2 (age) × 2 (group: stop vs. control) ANOVA on the predicted number of recalled words did not yield a significant main effect of age, $F < 1$, $\eta_p^2 = .001$, nor a main effect of group, $F < 1$, $\eta_p^2 = .001$, as well as no interaction, $F(1, 122) = 1.02$, $MSE = 73.92$, $p = .315$, $\eta_p^2 = .008$. Thus, the predicted number of recalled words was comparable for young ($M = 20.22$, $SD = 7.98$) and older adults ($M = 20.71$, $SD = 9.14$), and comparable for participants in the control group ($M = 20.14$, $SD = 8.51$) and the stop group ($M = 20.79$, $SD = 8.63$). Furthermore, in the stop group, the correlations between predicted 'stop' positions and the predicted number of recalled words were not significant for either young adults, $r(32) = -.10$, $p = .606$, or older adults, $r(31) = .12$, $p = .510$.

Obviously, comparing predicted recall with actual recall in Experiment 1 (List 1) revealed that participants' predictions grossly exceeded actual memory performance (main effect of experiment: $F(1, 246) = 205.07$, $MSE = 46.13$, $p < .001$, $\eta_p^2 = .455$), especially in case of older adults (Age × Experiment interaction: $F(1, 246) = 9.74$, $MSE = 46.13$, $p = .002$, $\eta_p^2 = .038$).

In Experiment 3b, a 3 (age: young, middle-aged, older) × 2 (condition: stop vs. control) mixed ANOVA on the predicted number of recalled words did not yield a significant main effect of age, $F(2, 87) = 1.55$, $MSE = 114.48$, $p = .218$, $\eta_p^2 = .034$, nor a main effect of condition, $F < 1$, $\eta_p^2 = .001$, and no interaction, $F(2, 87) = 2.67$, $MSE = 9.53$, $p = .075$, $\eta_p^2 =$

.058. Because the interaction was close to statistical significance, we compared participants' predicted recall performance separately for each age group. Pairwise comparisons (with the Bonferroni adjustment) showed that young participants predicted that they would have recalled more words in the stop condition (M = 14.70, SD = 8.05) than in the control condition (M = 13.04, SD = 7.42), a difference of 1.67, SE = 0.84, p = .050. For the middle-aged and older adults, this difference was not significant: -0.41, SE = 0.81, p = .611, and -0.81, SE = 0.75, p = .283, respectively.

Correlations between predicted 'stop' positions and the predicted numbers of words recalled were not significant either for young adults, $r(27) = .10$, $p = .625$, or for middle-aged adults, $r(29) = .18$, $p = .344$. However, this correlation was significant for older adults, $r(34) = .83$, $p < .001$.

Finally, for young and older groups only, we compared predicted recall performance with correctly recalled words in the stop condition of Experiment 1 (List 1). Once again, participants' predictions far exceeded actual performance (main effect of experiment: $F(1, 121) = 52.11$, $MSE = 39.39$, $p < .001$, $\eta_p^2 = .301$), especially in the case of older adults (Age × Experiment interaction: $F(1, 121) = 5.66$, $MSE = 39.39$, $p = .019$, $\eta_p^2 = .045$).

Discussion

In all age groups, participants chose comparable 'stopping' positions. Young and older adults' willingness to stop as well as the forecasted positions corresponded to those observed in Experiment 1. This pattern of results provides further support for the task perception hypothesis – it stresses the significant contribution of participants' assumptions about what is the most advantageous learning strategy in deciding to stop learning. However, when we inspected participants' responses to Experiment 3b's post-experiment questions (see Supplementary Materials), a frequent motive given was overload prevention. Moreover, combined results from both Experiments 3a and 3b indicated that more older than young

adults would decide to stop learning. We return to this apparent contradiction in the General Discussion.

Additionally, participants were remarkably optimistic in their predictions of their memory performance, especially older adults. These overestimations were particularly striking (relative to actual performance) in the stop condition (cf. Murayama et al., 2016, Experiment 5), indicating a widespread lack of awareness of the negative consequences of learning termination. Indeed, in Experiment 3b's questionnaire, the majority of participants declared that stopping learning is a beneficial strategy to remember more words. Recall improvement was also a popular answer to the open-ended questions about the reasons for stopping (see Supplementary Materials).

General Discussion

An adroit learner is capable not only of remembering a lot of information but also successful management and adjustment of their learning process to obtain desirable outcomes (Fiechter, Benjamin, & Unsworth, 2016). Thus, to explain differences in memory performance between two people, we need to know not only their memory capacity but also their metamemory abilities in a given task – namely, how proficient they are in monitoring and controlling their memory processes. Good metamemory skills can even help older adults to offset age-related declines in memory functioning (Hertzog, 2016). Besides that, a competent learner is sometimes able to obtain the same outcomes using less time and effort in comparison to a person having the same memory abilities but not so skilled in choosing the optimal learning strategies (Pyc & Rawson, 2007). Most often, however, effective learning demands overcoming a natural human tendency to avoid cognitive effort (Botvinick & Rosen, 2009). Our results, taken together, show that when young and older adults can decide whether to stop learning, the majority of them do so, even though it would be better for their performance not to. Moreover, the position at which participants stopped was positively

correlated with their memory performance. These observations replicate and noticeably broaden the results of Murayama et al.'s (2016) experiments in which only young adults took part.

The main aim of our research was to compare how young and older adults decide to stop studying to-be-remembered material before it is presented in its entirety, and what consequences this decision has for memory retention. We also examined participants' assumptions about learning termination in Experiments 3a and 3b. The results generally gave a coherent picture: a similar proportion of young and older adults decided to stop (Experiments 1 and 2), they did so at a comparable position (Experiment 1), and these decisions decreased the number of correctly recalled words equally in the two age groups (Experiments 1 and 2). Participants' predictions in Experiments 3a and 3b corresponded to actual learning decisions observed in Experiment 1 in terms of the number of participants who chose the 'stop' option as well as their 'stopping' positions. However, young and older adults predicted that they would recall a greater number of words than participants actually did in Experiment 1. Older adults were particularly overconfident. Even though their predictions were similar to young adults, their actual memory performance was worse in both Experiments 1 and 2. Crucially, in Experiment 3a, both young and older adults predicted that they would recall a similar number of words in the stop and control scenarios. In Experiment 3b, these results were replicated in the middle-aged and older adult groups. Young adults declared that they would remember an even greater number of words in the stop scenario. Thus, regardless of age, participants seem to be unaware of the negative consequences of learning termination in a self-regulated task.

Murayama et al. (2016) argued that "mental disfluency due to memory capacity overload is a plausible factor that influenced participants' decisions to stop receiving further information" (p. 921). These authors additionally mentioned an alternative explanation – the

role of participants' beliefs and assumptions about the task that may prompt them to make a 'stop' decision – and suggested that "experiments that include a short post-experiment survey asking for their strategies or intentions would clarify the nature of participants' beliefs" (p. 921). In our experiments, we followed up these ideas. We argued that the comparison of young and older adults might help in discrimination between two alternative hypotheses: cognitive overload and task perception. Our results mostly supported the latter. Although older adults may be more prone to being overwhelmed during learning, they neither terminated learning more frequently than young adults (Experiments 1 and 2) nor stopped at earlier positions (Experiment 1). As a result, even though their memory performance proved to be worse than young adults, they suffered comparable negative consequences of the 'stop' decisions. We also tested participants in a situation where any feeling of being overloaded was avoided (Experiments 3a and 3b), and participants had to decide on the basis of the task description alone (thus, having only their assumptions about the task at their disposal). Young and older adults' predictions corresponded to the results observed during actual learning episodes.

However, three caveats should be noted here. Firstly, simultaneous operations of both experience-based and theory-based metamemory judgments in steering control decisions during learning are not precluded (Koriat, 1997). Indeed, the feeling of being overloaded may have played a particular role, for example, in the tendency of older adults to stop a list earlier than young adults in Experiment 2. Hence, future research should manipulate encoding fluency directly. Secondly, the fact that participants' assumptions about the task played a significant role does not mean that participants need to have exactly the same beliefs about beneficial learning strategies. Indeed, in answering open-ended questions about reasons for stopping, older adults frequently mentioned their poor memory abilities. This reason was not observed in other age groups (see Supplementary Materials). Moreover, if non-adaptive

beliefs about the task motivated people to terminate learning prematurely, future experiments should include conditions in which participants are informed about the consequences of their self-regulated learning decisions. For example, experimental instructions may contain a warning notice: “research shows that stopping the learning of a word list early decreases the number of recalled words”. With such an intervention, researchers can check whether participants continue using the detrimental memory strategy even though their beliefs about the task have been corrected.[11] Thirdly, it is also theoretically possible, although not confirmed by the results of our studies, that young and older adults share similar experience-based and theory-based metamemory judgements but exercise control over to-be-remembered materials differently. Therefore, further research should explore experimental manipulations that examine in the ‘stop’ paradigm participants’ monitoring and control processes separately.

Finally, there may appear to be some similarities between the ‘stop’ paradigm and research on the list-length effect. The list-length effect is frequently associated with the observation that the probability of recall or recognition is reduced when the number of to-be-remembered items increases. The longer a to-be-remembered list is, the more difficult it is to retrieve any particular item from the list, mostly because of intra-list interference (e.g., Shiffrin, 1970; Smith, 1979; but see Dennis & Chapman, 2010, for an exception). In studies on the list-length effect, researchers frequently measure the proportion of recalled words out of all encoded ones – a so-called input-bound measure (Koriat & Goldsmith, 1996). (This contrasts with output-bound accuracy, which can measure the correctness of memory report; see Supplementary Materials.) Others, however, have also reported absolute recall in list-length experiments (e.g., Badham, Whitney, Sanghera, & Maylor, 2017; Ward, 2002) and found that the number of correctly recalled items increases as list-length increases. We

[11] We thank an anonymous reviewer for this future research suggestion.

focused here on the number of correctly recalled words, not the proportion, for two reasons. First, the participants' task was to recall as many items from the entire list as they could. All participants potentially had the chance to learn 50 words; thus, the proportions should be calculated out of 50 items, instead of only the items encoded by a given person. For simplicity, we decided to report only the number of correctly recalled items, like Murayama et al. (2016) did. The second reason touches on the most crucial difference between these two research paradigms. Our experiments investigated the consequences of a control decision to stop learning to maximize the number of recalled words in the test (i.e., a *metamemory* process). In contrast, research on the list-length effect examines how the amount of to-be-remembered material determines the probability of retrieval of a particular item (i.e., a *memory* process).

To conclude, being able to learn new information efficiently is crucial at all stages of life: from the earliest years and well into older age. Yet we all know from experience that we are often confronted with situations in which the amount of information we have to remember is likely to exceed our memory capability. Together, our results show that when young and older adults can decide whether to stop learning, the majority of them do so, even though their performance would have benefited if they had not. Importantly, young and older adults tend to stop learning at a similar point and, thus, they suffer comparable consequences of such a decision. Our data, therefore, suggest that beliefs about the task are more critical for the decision to stop than the feeling of cognitive overload. Moreover, our results indicate that, in some tasks, young and older adults are equally able to exert metamemory control, even though their decisions may not be beneficial for memory performance. These findings enhance our understanding of fundamental learning decisions and contribute to addressing "when and why older adults show similar or different patterns of metacognitive control"

(Hertzog, 2016, p. 554). As a result, this novel understanding could be translated into specific guidelines on how young and older adults can learn more effectively.

References

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*, 445–459. doi: 10.3758/BF03193014

Badham, S. P., Whitney, C., Sanghera, S., & Maylor, E. A. (2017). Word frequency influences on the list length effect and associative memory in young and older adults. *Memory, 25*, 816–830. doi: 10.1080/09658211.2016.1224358

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

Bjork, R. A. (2003). Interference and forgetting. In J. H. Byrne (Ed.), *Learning and memory* (pp. 268–273). Macmillan Reference USA: Thomson/Gale.

Bjork, E. L., de Winstanley, P. A., & Storm, B. C. (2007). Learning how to learn: can experiencing the outcome of different encoding strategies enhance subsequent encoding? *Psychonomic Bulletin & Review*, *14*, 207–211. doi: 10.3758/BF03194053

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

Botvinick, M. M., & Rosen, Z. B. (2009). Anticipation of cognitive demand during decision-making. *Psychological Research, 73*, 835–842. doi: 10.1007/s00426-008-0197-8

Brehm, J. W. (1966). *A theory of psychological reactance.* Academic Press.

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the life span: a cross-sectional study of memory and selectivity. *Developmental Psychology*, *47*, 1553–1564. doi: 10.1037/a0025623

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: age-related differences and similarities in self-regulated learning. *Psychology and Aging, 28*, 232–242. doi: 10.1037/a0030678

Coane, J. H., & Umanath, S. (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. *Journal of Memory and Language, 107*, 152–168. doi: 10.1016/j.jml.2019.05.002

Dennis, S., & Chapman, A. (2010). The inverse list length effect: A challenge for pure exemplar models of recognition memory. *Journal of Memory and Language, 63*, 416–424. doi: 10.1016/j.jml.2010.06.001

Dobbins, I. G., & Kantner, J. (2019). The language of accurate recognition memory. *Cognition, 192*, 103988. doi: 10.1016/j.cognition.2019.05.025

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: differential and sufficient aspects. *Journal of Experimental Child Psychology, 47*, 274–296. doi: 10.1016/0022-0965(89)90033-7

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 54. The psychology of learning and motivation: Advances in research and theory* (pp. 103–140). Elsevier Academic Press.

Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky & R. A. Bjork (Eds.), *A handbook of metamemory and memory* (pp.11-28). Psychology Press.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29*, 247–253. doi: 10.3758/bf03194918

Fiechter, J. L., Benjamin, A. S., & Unsworth, N. (2016). The metacognitive foundations of effective remembering. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 307–324). Oxford University Press. doi: 0.1093/oxfordhb/9780199336746.013.24

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 632–652. doi: 10.1037/a0026215

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 193–225). Academic Press. doi: 10.1016/S0079-7421(08)60041-9

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance. Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). The MIT Press.

Healey, M. K., Campbell, K. L., & Hasher, L. (2008). Cognitive aging and increased distractibility: costs and potential benefits. *Progress in Brain Research*, *169*, 353–363. doi: 10.1016/S0079-6123(07)00022-2.

Hertzog, C. (2016). Aging and metacognitive control. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 537–558). Oxford University Press. doi: 10.1093/oxfordhb/9780199336746.013.31

Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. (2003). Latent change models of adult cognition: are changes in processing speed and working memory

associated with changes in episodic memory? *Psychology and Aging*, *18*, 755–769. doi: 10.1037/0882-7974.18.4.755

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 417–466). Lawrence Erlbaum Associates Publishers.

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., Small, B. J., McFall, G. P., & Dixon, R. A. (2019). Age, cohort, and period effects on metamemory beliefs. *Psychology and Aging*, *34*, 1077–1089. doi: 10.1037/pag0000384

Hess, T. M. (2014). Selective engagement of cognitive resources: Motivational influences on older adults' cognitive functioning. *Perspectives on Psychological Science*, *9*, 388–407. doi: 10.1177/1745691614527465

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

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

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., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656. doi: 10.1037/0096-3445.133.4.643

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517. doi: 10.1037/0033-295x.103.3.490

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: the benefits—and costs—of dropping flashcards. *Memory*, *16*, 125–136. doi: 10.1080/09658210701763899

Kornell, N., & Finn, B. (2016). Self-regulated learning: an overview of theory and data. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 325–340). Oxford University Press. doi: 10.1093/oxfordhb/9780199336746.013.23

Kuhlmann, B. G., & Undorf, M. (2018). Is all metamemory monitoring spared from aging? A dual-process examination. *Psychology and Aging*, *33*, 1152–1167. doi: 10.1037/pag0000318.

Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica*, *36*, 280–289. doi: 10.1016/0001-6918(72)90012-1

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). Subtlex-pl: subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, *47*, 471–483. doi: 10.3758/s13428-014-0489-4.

McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement by younger and older adults. *Psychology and Aging*, *26*, 137–142. doi: 10.1037/a0022681

Murayama, K., Blake, A. B., Kerr, T., & Castel, A. D. (2016). When enough is not enough: information overload and metacognitive decisions to stop studying information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 914–924. doi: 10.1037/xlm0000213

Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). Academic Press.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-25). MIT Press.

Old, S. R., & Naveh-Benjamin, M. (2008). Memory for people and their actions: further evidence for an age-related associative deficit. *Psychology and Aging*, *23*, 467–472. doi: 10.1037/0882-7974.23.2.467

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

Price, J., Hertzog, C., & Dunlosky, J. (2010). Self-regulated learning in younger and older adults: does aging affect metacognitive control? *Aging, Neuropsychology, and Cognition*, *17*, 329–359. doi: 10.1080/13825580903287941

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*, 1917–1927. doi: 10.3758/BF03192925

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 5: The Mill Hill Vocabulary Scale.* Harcourt Assessment.

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

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428. doi: 10.1037/0033-295x.103.3.403

Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, *34*, 17–24. doi: 10.1037/pag0000288

Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling-of-knowing. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *39*, 467–475. doi: 10.1037/h0080074

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, *168*, 1601–1603. doi: 10.1126/science.168.3939.1601

Smith, A. D. (1979). The interaction between age and list length in free recall. *Journal of Gerontology*, *34*, 381–387. doi: 10.1093/geronj/34.3.381

Storm, B. C., Hickman, M. L., & Bjork, E. L. (2016). Improving encoding strategies as a function of test knowledge and experience. *Memory & Cognition*, *44*, 660–670. doi: 10.3758/s13421-016-0588-9.

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, *43*, 27–38. doi:10.3758/s13421-014-0455-5.

Ward, G. (2002). A recency-based account of the list length effect in free recall. *Memory & Cognition*, *30*, 885–892. doi: 10.3758/bf03195774

Wechsler, D. (1981). *Manual for the Wechsler adult intelligence scale – revised*. Psychological Corporation.

Table 1

Numbers (Out of 32) and Percentages of Young and Older Adults who Decided to Stop Learning in Experiments 1 and 2

	Young						Older					
	List 1		List 2		List 3		List 1		List 2		List 3	
Experiment	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Exp. 1	19	59.4	19	59.4	18	56.3	20	62.5	22	68.8	23	71.9
Exp. 2: stop-first	23	71.9	24	75.0	---	---	21	65.6	23	71.9	---	---
Exp. 2: stop-second	22	68.8	21	65.6	---	---	26	81.3	25	78.1	---	---

Table 2

Mean Positions (and Standard Deviations) at Which Young and Older Adults Decided to Stop Learning in Experiments 1 and 2

	Young			Older		
	List 1	List 2	List 3	List 1	List 2	List 3
Experiment	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Exp. 1	29.63 (19.27)	30.75 (17.96)	33.59 (16.53)	34.03 (14.90)	28.41 (16.84)	26.81 (16.45)
Exp. 2: stop-first	35.25 (14.43)	30.09 (15.46)	---	31.84 (16.00)	25.97 (17.52)	---
Exp. 2: stop-second	33.53 (15.29)	32.66 (16.06)	---	24.72 (15.79)	23.91 (16.42)	---

Table 3

Numbers and Percentages of Participants who Predicted They Would Decide to Stop Learning, and Their Mean (and Standard Deviation) Predicted Stopping Positions, in Experiments 3a and 3b

Variable		Young	Middle-aged	Older
Stop decision		N (%)	N (%)	N (%)
	Exp. 3a	15/32 (46.9)	---	22/31 (71.0)
	Exp. 3b	12/27 (44.4)	18/29 (62.1)	21/34 (61.8)
Stop position		*M (SD)*	*M (SD)*	*M (SD)*
	Exp. 3a	35.53 (16.63)	---	30.74 (15.30)
	Exp. 3b	37.87 (14.59)	31.69 (15.43)	32.71 (16.05)

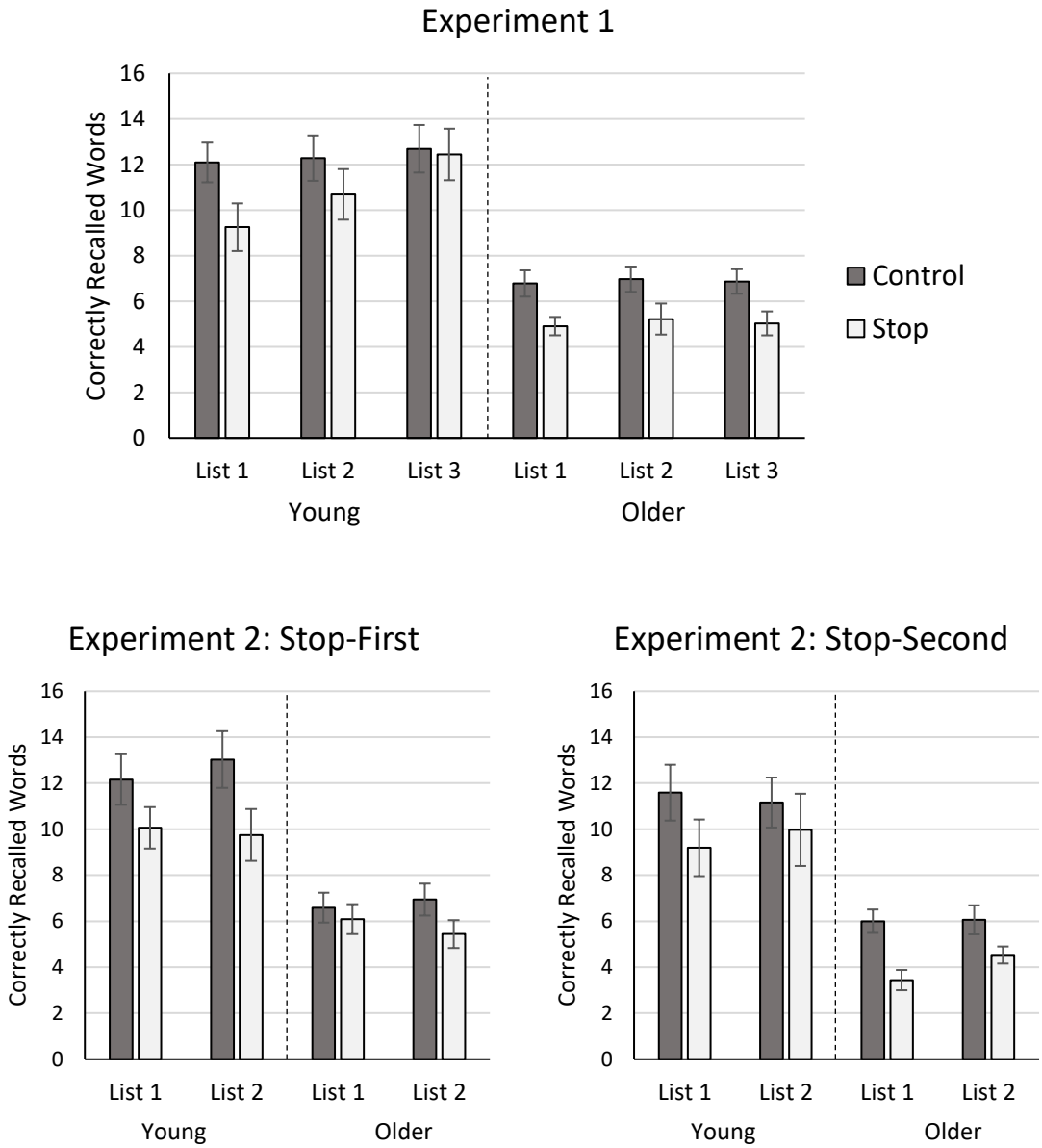

Figure 1. Mean numbers of words correctly recalled by young and older adults in Experiments 1 and 2. Error bars represent standard errors.

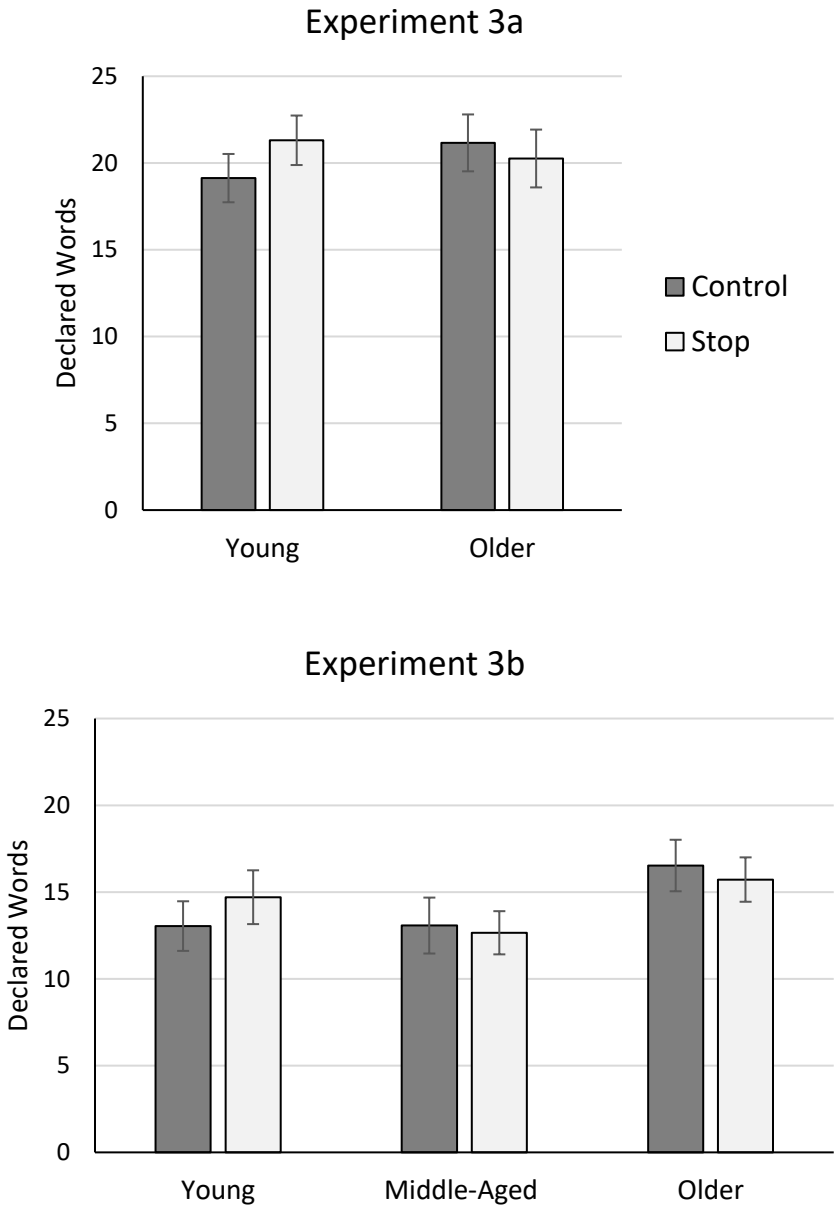

Figure 2. Mean numbers of words declared as recall predictions by young and older participants in Experiment 3a, and by young, middle-aged, and older participants in Experiment 3b. Error bars represent standard errors.

Appendix

Table A1

Mean Numbers (and Standard Deviations) of Recalled Words (Total, Correct, and Incorrect) by Young and Older Adults in Experiments 1 and 2

	Young						Older					
	Total		Correct		Incorrect		Total		Correct		Incorrect	
	Control	Stop	Control	Stop	Control	Stop	Control	Stop	Control	Stop	Control	Stop
Experiment	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Exp.1												
List 1	13.41 (4.67)	9.88 (5.80)	12.09 (4.93)	9.25 (5.92)	1.31 (1.26)	0.63 (0.94)	8.03 (3.15)	5.81 (2.24)	6.78 (3.26)	4.91 (2.28)	1.25 (1.05)	0.91 (1.06)
List 2	13.72 (6.07)	11.50 (6.22)	12.28 (5.63)	10.69 (6.28)	1.44 (1.29)	0.81 (0.82)	8.22 (2.86)	6.19 (4.27)	6.97 (3.14)	5.22 (3.86)	1.25 (1.30)	0.97 (1.09)
List 3	14.22 (5.91)	13.31 (6.66)	12.69 (5.89)	12.44 (6.39)	1.53 (1.69)	0.88 (1.21)	8.41 (3.35)	6.16 (3.00)	6.88 (3.05)	5.03 (2.98)	1.53 (1.63)	1.13 (1.29)
Exp. 2: stop-first												
List 1	13.41 (6.21)	10.81 (5.33)	12.16 (6.21)	10.06 (5.09)	1.25 (1.70)	0.75 (0.95)	8.00 (4.01)	7.03 (4.04)	6.59 (3.67)	6.09 (3.68)	1.41 (1.32)	0.94 (0.98)
List 2	14.72 (7.19)	10.47 (6.34)	13.03 (6.97)	9.75 (6.37)	1.69 (1.80)	0.72 (0.77)	8.97 (4.63)	6.34 (3.59)	6.94 (3.93)	5.44 (3.43)	2.03 (1.99)	0.91 (1.00)
Exp. 2: stop-second												
List 1	12.66 (6.69)	10.66 (7.05)	11.59 (6.87)	9.19 (6.97)	1.06 (1.61)	1.47 (1.92)	7.38 (2.88)	4.53 (2.74)	6.00 (2.87)	3.44 (2.49)	1.38 (1.41)	1.09 (1.23)
List 2	12.28 (5.84)	11.16 (8.48)	11.16 (6.14)	9.97 (8.89)	1.13 (1.19)	1.19 (1.33)	7.63 (3.47)	5.56 (2.59)	6.06 (3.57)	4.53 (2.09)	1.56 (1.59)	1.03 (1.28)

Supplementary Materials

Experiments 1 and 2

We analyzed participants' response accuracy by dividing the number of correctly recalled words by all volunteered responses in the memory test. This measure is frequently labelled as *output-bound accuracy,* and it refers to the degree of correctness of the information volunteered in free-report memory tests (Koriat & Goldsmith, 1996).[12] In their analysis, Maruyama and colleagues (Murayama, Blake, Kerr, & Castel, 2016) focused on the total number of correctly recalled words. However, we suspected that, although participants were explicitly instructed that their goal in our tasks was to recall as many words as possible, they may nonetheless be partially driven by the willingness to provide only correct responses in the memory test. Simply, they might have treated 'stop' decisions as a helpful strategy to maintain high accuracy of their responses, according to the following reasoning: 'I will learn fewer words, but I will remember them better.' Thus, we decided to introduce an accuracy measurement as an enhancement of our quantity measurement expressed in numbers of correctly recalled words.

Experiment 1 – Memory Accuracy

One older adult in the stop condition was excluded from this analysis as they did not correctly recall any word from one of the to-be-remembered lists. Figure S1 illustrates the mean output-bound accuracy of young and older participants, separately for each experimental group. The proportion of correctly recalled words out of all words provided in the memory tests was calculated (i.e., accuracy) as a function of age (young vs. older), list (1,

[12] An alternative measure is the *input-bound accuracy,* which concerns the amount of information provided in the memory test; it refers to the proportion of correctly recalled words out of all words encoded during the study phase. However, this proportion could not be consistently applied to our study. The reason is that the amount of to-be-encoded materials was not equal for all participants (in the stop conditions, participants could decide to stop learning at a point of their choosing). Therefore, in Experiments 1 and 2, we decided to report the number of correctly recalled words as a measure of the quantity of memory report, inspired by the study of Murayama et al. (2016).

2, 3) and experimental group (stop vs. control). A 2 (age) × 3 (list) × 2 (group) mixed ANOVA on accuracy yielded a significant main effect of age, $F(1, 123) = 12.24$, $MSE = 0.05$, $p < .001$, $\eta_p^2 = .091$. Young adults were more accurate in the memory tests ($M = .90$, $SD = .09$) than were older adults ($M = .83$, $SD = .15$). The remaining main effects and interactions were not significant. Crucially, the main effect of group did not reach significance, $F < 1$, $\eta_p^2 = .003$, indicating that participants in the stop group ($M = .87$, $SD = .13$) were no more accurate than those in the control group ($M = .86$, $SD = .12$). Thus, if participants terminated learning as a strategy to improve response accuracy in this experiment, it was not successful.

Experiment 2 – Memory Accuracy

Three older adults (one participant from the stop-first group, two from the stop-second group) were removed from this analysis as they did not correctly recall any word from one of the to-be-remembered lists. Figure S1 illustrates the mean output-bound accuracy of young and older participants. Accuracy was analyzed as a function of participants' age (young vs. older), condition (stop vs. control), list (1 vs. 2), and experimental group (stop-first vs. stop-second) using a four-factor mixed ANOVA. This yielded a significant effect of age, $F(1, 121) = 16.51$, $MSE = 0.04$, $p < .001$, $\eta_p^2 = .120$, indicating again that young adults ($M = .88$, $SD = .09$) were significantly more accurate than were older adults ($M = .82$, $SD = .10$). The main effect of group was also significant, $F(1, 121) = 4.13$, $MSE = 0.04$, $p = .044$, $\eta_p^2 = .033$. This effect was qualified by a significant interaction between group and condition, $F(1, 121) = 4.96$, $MSE = 0.03$, $p = .028$, $\eta_p^2 = .039$, which arose simply because participants became less accurate from the first two lists (.87) to the second two lists (.83). None of the remaining main effects and interactions reached statistical significance.

Experiment 3b

In Experiment 3b, participants predicted whether they would stop learning solely on the basis of reading a hypothetical scenario. To explore potential reasons for stopping learning, after making their predictions, participants were asked to provide a written answer to an open-ended question, either “In the stop scenario, if you chose to stop, why?” or “If you chose not to stop, why not?”. Finally, they indicated whether or not they agreed with each of four statements relating to possible mechanisms derived from the self-regulated learning literature. We now summarize responses to each of these sets of questions in turn.

Open-ended Responses

Here, we were interested in how participants justified their metacognitive decisions in plain English (see Coane & Umanath, 2019; Dobbins & Kantner, 2019; for examples of similar procedures). We excluded from further analysis answers either when participants refrained from providing any specific reason (N = 5; e.g., *Bored with same questions*) or it seemed that they misunderstood the experimental procedure after having read the scenario (N = 2; e.g., *Slowing down may allow earlier words to be forgotten*). For each decision, we identified five main reasons. Table S1 shows all these reasons with their definitions and examples. Each participant’s response received a score of 1 indicating that a given reason was present, or 0 when a reason was absent. Thus, when a participant offered more than one reason, a single response could be coded more than once. Two independent coders (authors AK and AJ) individually assessed participants’ answers; correlations in their evaluations ranged from .61 to 1. They then discussed all discrepancies in their assessment and agreed on the final data set. The data were analyzed for each question separately.

Reasons to stop learning. In total, 10 young, 18 middle-aged, and 21 older adults gave valid answers to the question about reasons for stopping learning. Means and standard deviations are presented in Table S2. We performed a 3 (age) × 6 (reasons: increase recall, information overload, gain of control, memory strategy, alleviating pressure, poor memory)

mixed ANOVA, which produced a significant effect of reason, $F(4, 184) = 5.62$, $MSE = 0.19$, $p < .001$, $\eta_p^2 = .109$, as well as an interaction, $F(8, 184) = 2.49$, $MSE = 0.19$, $p = .014$, $\eta_p^2 = .098$. The main effect of age was not significant, $F < 1$, $\eta_p^2 = .011$. In general, two reasons dominated in participants' answers: motivation to increase recall and tendency to avoid a feeling of being overwhelmed by the amount of to-be-remembered materials. One-way ANOVAs on each reason separately suggested that the interaction was attributable to young adults tending to report information overload as their reason for stopping more often than middle-aged and older adults; middle-aged adults reporting a memory strategy more than older adults; and older adults reporting poor memory as their reason more often than young and middle-aged adults.

Reasons not to stop learning. In total, 14 young, 10 middle-aged, and 12 older adults gave valid answers to the question about reasons in favor of trying to encode the entire to-be-remembered material. A 3 (age) × 6 (reasons: increase recall, challenge, completeness, memory strategy, self-testing, concentration) mixed ANOVA did not yield a significant effect of age, $F < 1$, $\eta_p^2 = .049$, a significant effect of reason, $F(4, 132) = 1.29$, $MSE = 0.21$, $p = .279$, $\eta_p^2 = .038$, nor an interaction, $F < 1$, $\eta_p^2 = .030$. Nonetheless, numerically the most common reason overall (and for young and older adults in particular) for learning the entire to-be-remembered material was that it would increase their recall, while the second most common reason overall (and the most popular for middle-aged adults) was a determination to complete the learning. Finally, like the reasons for stopping, older adults were the least likely age group (at least numerically) to report an incorporation of other memory strategies.

Forced-choice Responses

The effects of age in answering the forced-choice questions (see Figure S2) were assessed via four separate χ^2 tests to investigate whether young, middle-aged, and older adults evaluated each statement similarly. Overall, a similar number of participants in each age

group agreed that ending the list early (a) ‘is a good strategy to help recall as many words as possible’, $\chi^2(2) = 0.48$, $p = .787$; (b) ‘will prevent overloading of memory’, $\chi^2(2) = 0.03$, $p = .985$; (c) ‘will help ensure that all the responses are correct’, $\chi^2(2) = 0.89$, $p = .641$, and (d) ‘will reduce the effort in learning the words’, $\chi^2(2) = 1.89$, $p = .388$.

After collapsing the data across age groups, one Cochran’s Q test for all participants was employed to assess whether there was a difference in the degree of agreement with the statements included in the post-experimental questionnaire. Participants did not agree with these statements equally: $\chi^2(3) = 20.94$, $p < .001$ (see Figure S2). An exact McNemar’s test revealed significant differences between the third statement and each of the other statements (the first, $p = .020$; the second, $p < .001$; the fourth, $p = .002$), and also between the first and second statements ($p = .049$).

References

Coane, J. H., & Umanath, S. (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. *Journal of Memory and Language*, *107*, 152–168. doi: 10.1016/j.jml.2019.05.002

Dobbins, I. G., & Kantner, J. (2019). The language of accurate recognition memory. *Cognition*, *192*, 103988. doi: 10.1016/j.cognition.2019.05.025

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517. doi: 10.1037/0033-295x.103.3.490

Murayama, K., Blake, A. B., Kerr, T., & Castel, A. D. (2016). When enough is not enough: Information overload and metacognitive decisions to stop studying information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 914–924. doi: 10.1037/xlm0000213

Table S1

Reasons Used in Evaluations of Participants' Responses to Open-Ended Questions After Reading the Stop Scenario in Experiment 3b

Reason	Definition	Sample Answers
	Question: If you chose to stop, why?	
increase recall	Responses indicated that learning termination may help in remembering more words.	*More chance of remembering more words*
information overload	Responses mentioned a high number of to-be-remembered words and/or a feeling of being overloaded during learning.	*Too many to remember.*
memory strategy	Responses mentioned other learning strategies than learning termination.	*I would try to picture the objects represented by the words interacting with each other or use other memory methods.*
alleviating pressure	Responses contained the word "pressure" or mentioned that learning termination is helpful in overcoming the tension felt when studying words.	*Feel less pressure.*
poor memory	Responses mentioned that a stop decision may be associated with poor memory abilities.	*I didn't think I could remember all the words.*
	Question: If you chose not to stop, why not?	
increase recall	Responses mentioned that trying to encode the entire list may help to remember more words.	*A higher number of words would likely mean a higher chance of remembering more.*
completeness	Responses stressed a reluctance to interrupt learning and/or willingness to see the entire list of words.	*Like to finish what I start.*
memory strategy	Responses mentioned other learning strategies than the decision to encode entire list of words.	*Try to associate/group words to make them easier to remember. Try to link as many as possible.*
self-testing	Responses mentioned that not stopping enables self-testing of memory abilities.	*To see how many I could remember*
concentration	Responses noted that not stopping may help to maintain concentration.	*Not to lose concentration.*

Table S2

Means Rates (and Standard Deviations) for Reasons Either in Favor or Against Stopping Learning

	Young	Middle-Aged	Older
Reasons	*M (SD)*	*M (SD)*	*M (SD)*
	Question: If you chose to stop, why?		
increase recall	.20 (.42)	.39 (.50)	.33 (.48)
information overload	.70 (.48)	.33 (.49)	.29 (.46)
memory strategy	.20 (.42)	.39 (.50)	.05 (.22)
alleviating pressure	.00 (.00)	.06 (.24)	.14 (.36)
poor memory	.00 (.00)	.00 (.00)	.29 (.46)
	Question: If you chose not to stop, why not?		
increase recall	.36 (.50)	.20 (.42)	.33 (.49)
completeness	.21 (.43)	.40 (.52)	.25 (.45)
memory strategy	.29 (.47)	.30 (.48)	.08 (.29)
self-testing	.21 (.43)	.20 (.42)	.17 (.39)
concentration	.07 (.27)	.00 (.00)	.17 (.39)

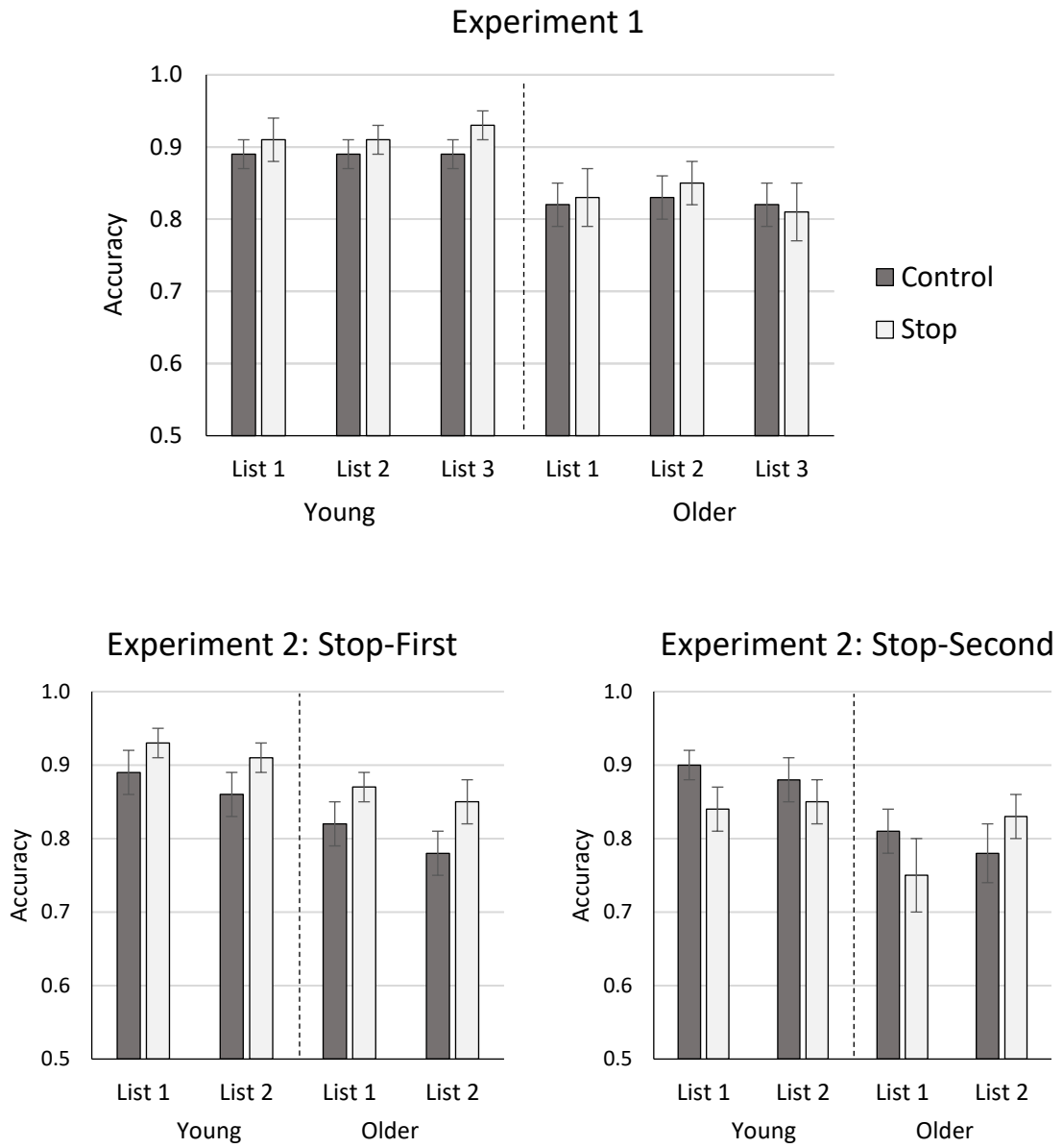

Figure S1. Young and older participants' output-bound accuracy in Experiments 1 and 2. Error bars represent standard errors.

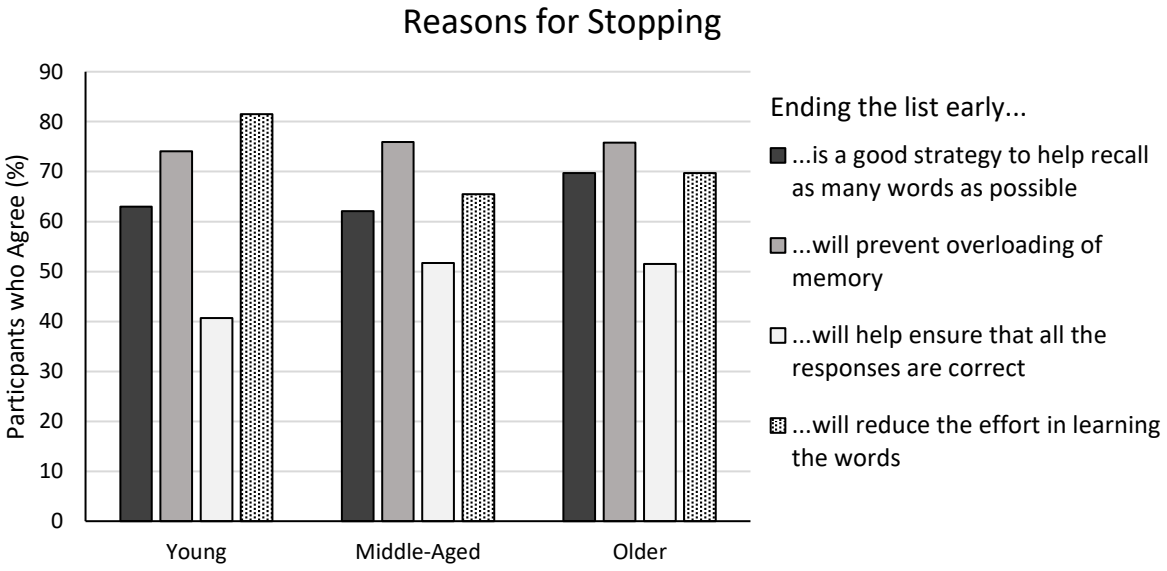

Figure S2. Percentages of participants who agreed with each statement in the post-experimental questionnaire in Experiment 3b.