

On Teaching Old Dogs New Tricks

Psychological Science
1–10

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797615597912

pss.sagepub.com



Janet Metcalfe¹, Lindsey Casal-Roscum², Arielle Radin^{1,2}, and David Friedman^{1,2,3}

¹Department of Psychology, Columbia University; ²Division of Cognitive Neuroscience, New York State Psychiatric Institute, New York, New York; and ³Department of Psychiatry, Columbia University

Abstract

Although older adults rarely outperform young adults on learning tasks, in the study reported here they surpassed their younger counterparts not only by answering more semantic-memory general-information questions correctly, but also by better correcting their mistakes. While both young and older adults exhibited a *hypercorrection effect*, correcting their high-confidence errors more than their low-confidence errors, the effect was larger for young adults. Whereas older adults corrected high-confidence errors to the same extent as did young adults, they outdid the young in also correcting their low-confidence errors. Their event-related potentials point to an attentional explanation: Both groups showed a strong attention-related P3a in conjunction with high-confidence-error feedback, but the older adults also showed strong P3as to low-confidence-error feedback. Indeed, the older adults were able to rally their attentional resources to learn the true answers regardless of their original confidence in the errors and regardless of their familiarity with the answers.

Keywords

aging, memory, evoked potentials

Received 11/5/14; Revision accepted 7/6/15

Older adults rarely outperform young adults on cognitive tasks (Balota, Dolan, & Duchek, 2000). But there are a few exceptions. One well-documented instance is older adults' performance on semantic memory tasks (Staudinger, Cornelius, & Baltes, 1989; Umanath & Marsh, 2014). For example, when asked to provide the answers to general-information questions, healthy older adults outperform young adults (McIntyre & Craik, 1987; Perlmutter, 1978). The dominant explanation for older adults' superior semantic memory performance is that, by virtue of having lived longer, they have a larger store of semantic knowledge. Additionally, the knowledge they have accumulated has become rigid, or "crystallized," and hence is thought to be shielded from overwriting. A larger store of old information coupled with a generalized difficulty with new learning, which protects and renders old semantic knowledge relatively impervious to change, is forwarded as an explanation for this superior semantic performance (Botwinick, 1984; Jacoby, Hessels, & Bopp, 2001).

The present research challenges the view that such crystallization is endemic to normal aging. Recent studies

of semantic-error correction (Eich, Stern, & Metcalfe, 2013; see Cyr & Anderson, 2013; Sitzman, Rhodes, Tauber, & Licalde, 2015) suggest the possibility, which we investigated in the study reported here, that older healthy adults not only are better than young adults at answering general-information questions in the first place, but also, when they do make a mistake, might be more likely than young adults to correct those errors. Correcting errors is, of course, the quintessential new-learning task: To correct mistakes, one needs to supplant entrenched responses with new ones. If older adults display greater facility at error correction than young adults, this would directly contravene the view that aging necessarily produces cognitive rigidity and an inability to learn.

In the error-correction studies, participants were given a series of general-information questions, such as "What

Corresponding Author:

David Friedman, Columbia University, Department of Psychiatry, Cognitive Electrophysiology Laboratory – Unit 50, Division of Cognitive Neuroscience, 1051 Riverside Dr., New York, NY 10032
E-mail: df12@cumc.columbia.edu

was the last name of the woman who founded the American Red Cross?" After answering each question, participants indicated their confidence in the correctness of their answer. They were then given the correct answer. This continued, in the Eich et al. (2013) study, for example, until participants had committed 15 errors. Later, there was a surprise retest on these errors. Several results were of interest. First, the older adults had to be asked more questions before they reached the same number of incorrect responses as the young adults; that is, the older adults' semantic memory was better than that of young adults. Second, in the Eich et al. (2013) investigation, older adults were not different from younger adults in the proportion correct on the final retest on their previous errors.¹ Third, although the young adults corrected their high-confidence errors more than their low-confidence errors—an effect called the *hypercorrection effect* (e.g., Butler, Fazio, & Marsh, 2011; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009, 2010; Iwaki, Matsushima, & Kodaira, 2013; Kulhavy, Yekovich, & Dyer, 1976; Metcalfe, Butterfield, Habeck, & Stern, 2012; Metcalfe & Finn, 2012; Metcalfe & Miele, 2014; Sitzman et al., 2015)—older adults exhibited this difference to a much lesser extent (and see note 1). The smaller hypercorrection effect could be interpreted as a deficit in processing due to aging.

However, it is also possible that this pattern represents enhanced processing rather than a deficit. First, the fact that the older adults had to be asked more questions than the young adults to commit the same number of errors resulted in a pool of errors that was, on average, more difficult for the older participants. Equal performance on what is arguably a more difficult pool suggests that the older adults were *better* than the young adults at error correction (also see note 1). Second, the smaller hypercorrection effect among the older adults could have come about in two ways: (a) because the high-confidence errors were corrected *less* (which would suggest a deficit) or (b) because the low-confidence errors were corrected *more* (which would suggest enhanced learning).

To further investigate these differences, we conducted an event-related potential (ERP) study of error correction in young and older adults. One well-documented explanation of the hypercorrection effect is that the feedback to high-confidence errors attracts more attention than does the feedback to low-confidence errors (Butterfield & Mangels, 2003; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2010; cf. Metcalfe et al., 2012; Metcalfe & Finn, 2012). In a study consistent with the attentional view, Butterfield and Mangels (2003) showed that the feedback to high-confidence errors elicited an attention-related ERP component—the P3a—to a greater extent than did feedback to low-confidence errors, in young adults. The magnitude of the P3a component has been

associated both with attentional capture by salient events (Friedman, Cycowicz, & Gaeta, 2001; Ranganath & Rainer, 2003) and with memory encoding (Ranganath & Rainer, 2003). Moreover, some investigations have shown that a positive-going waveform, which includes the decision-related P3b component, is of greater amplitude for items that are later remembered compared with those that are forgotten (e.g., Fabiani & Donchin, 1995; Paller, Kutas, & Mayes, 1987; Ranganath & Rainer, 2003). Further, the hippocampus, a critical structure in memory, is also thought to be a part of the network that gives rise to the P3a (Friedman, Nessler, Kulik, & Hamberger, 2011; Knight, 1984).

Convergent with the idea that there are attentional differences accompanying the feedback to high- versus low-confidence errors in young adults, a functional MRI study (Metcalf et al., 2012) showed greater anterior cingulate cortex (ACC) activation associated with the feedback to high- as compared with feedback to low-confidence errors. Correspondingly, intracranial ERP recordings have demonstrated that the ACC is one of the generators of the attention-related P3a that is recorded at the scalp (Baudena, Halgren, Heit, & Clarke, 1995). Finally, behavioral studies also converge on an attentional factor in the hypercorrection effect. Butterfield and Metcalfe (2006) showed that when college students were asked to simultaneously detect soft tones while engaged in the error-correction task, they were likely to miss detecting the tone more during the presentation of corrective feedback to high-confidence errors than during the presentation of feedback to low-confidence errors. If the smaller hypercorrection effect in older adults is related to their failing to focus their attention on the feedback to high-confidence errors, they might be expected to show a decreased (or no) P3a to the high-confidence-error feedback.

However, the age-related difference in the hypercorrection effect could also have resulted not because the older adults paid less attention to their high-confidence errors than the young adults did but, rather, because they had superior processing on their low-confidence errors, paying more attention to them. This would show up as superior retest performance on the low-confidence errors, coupled with a "normal" P3a on the feedback to high-confidence errors and perhaps also an attentional P3a on the feedback to low-confidence errors.

Our hypotheses, then, were that (a) older adults would have greater semantic knowledge on the initial test than young adults; (b) when the difficulty of the error-correction task was equated by using the same set of items for both groups, older adults would correct their errors more often than young adults, showing superior new learning; (c) as in past research, older adults would show a smaller hypercorrection effect than young adults; (d) the smaller

hypercorrection effect among older adults would be due to better memory for the low-confidence-error feedback rather than worse memory for the high-confidence-error feedback; and (e) there would be no decrease in the P3a in response to high-confidence-error feedback in older adults, and there might be enhanced attentional processing (reflected by an enhanced P3a) in response to low-confidence-error feedback for these participants. According to the alternative *deficit* hypothesis, although semantic memory should be better for older adults than for young adults on the initial test, the older adults' error correction should be worse. It predicts that a smaller hypercorrection effect among older adults would be attributable to older adults failing to pay attention to the feedback to high-confidence errors, which should be reflected in little or no P3a voltage deflection in association with the high-confidence-error feedback.

Method

Participants

Forty-four young adults (25 women and 19 men) with a mean age of 24.20 years (range: 20–31 years) and a mean of 15.50 years of education ($SD = 1.60$) and 45 older adults (33 women and 12 men) with a mean age of 73.7 years (range: 62–88 years) and a mean of 16.8 years of education ($SD = 2.20$) completed the experiment and were paid to participate (for further details on the two groups, see Table 1). The data of 10 additional participants in each age group were eliminated because of excessive artifact and/or too few trials in the critical conditions. We did not have a strict advance criterion for the number of participants. However, because we were looking for a possibly subtle between-groups interaction on the magnitude of the P3a, we thought that we should approximately double the number of participants for each group from the number used in the only prior ERP investigation of hypercorrection (Butterfield & Mangels, 2003). Butterfield and Mangels recruited 25 young adults in their first experiment, and 20 provided usable data; in their second experiment, 23 participants were recruited, and 20 provided usable data. These experiments demonstrated a confidence-related effect on the P3a. Hence, we aimed for approximately 40 to 50 participants per age group to allow evaluation of a potential interaction of age group and confidence on the P3a.

All participants were native English speakers, with normal or corrected-to-normal vision and no history of neurological or psychiatric disorders; they were free from medications known to affect the central nervous system. Older adults were prescreened with an extensive telephone inventory, and responses were forwarded to a board-certified neurologist. The neurologist reviewed the

material and evaluated evidence for the presence of neurodegenerative disorders, neurovascular disease, and/or medications that might affect cognitive function. All older adult participants whose data are presented here were classified as aging normally. All participants provided informed consent according to the criteria of the New York State Psychiatric Institute's Institutional Review Board and were paid \$15 per hour.

Stimuli and procedures

The stimuli were 439 general-information questions on a variety of topics, taken from a set published by Nelson and Narens (1980), as well as various board games and Internet sites. All questions had answers that were single words 3 to 14 letters long (e.g., "In what ancient city were the Hanging Gardens located?" correct answer: Babylon).

The sequence for a single trial is presented in Table 2. The experiment comprised two phases, an initial test phase and a surprise retest phase. Electroencephalograms (EEGs) were recorded only during the first phase. General-information questions were presented in the center of the computer screen, and participants were given an unlimited amount of time to respond verbally. They were encouraged to guess if they were not sure of an answer, but were allowed to say, "I don't know" (i.e., to omit a response). The experimenter recorded the participants' responses by typing them on a keyboard. Next, except for omitted responses, participants used a keyboard to rate their confidence in the correctness of their response, on a scale ranging from 1 (*least confident*) to 7 (*most confident*). Participants were encouraged to use the entire scale. Questions were randomized separately for each participant.

Immediately following the confidence rating or omitted response, a central fixation point appeared. After 500 ms, a tone was delivered via speaker to announce the upcoming visual feedback, which occurred after a 1,500-ms delay. The feedback provided the correct answer. Following an error, the participant entered his or her familiarity with the correct answer (1 = familiar, 2 = unfamiliar).

There were 44 questions in each block, with a short break of approximately 5 min after each block. The experiment typically lasted 4 hr, because it took a long time to accumulate a sufficient number of high-confidence errors (we required at least 20), which are inherently rare. Because we set a minimum of five trials for inclusion in the ERP averages for each condition of the experiment, there was variability in the number of blocks administered to participants. Older participants viewed an average of 244 questions ($SD = 78$, range = 176–401), and young participants viewed an average of 230 questions ($SD = 66$, range = 132–396). The number of questions did not differ between young and older adults ($t < 1$).

Table 1. Demographic Data for the Young and Older Adults in the Study

Variable	Grand mean		Comparison of young and older adults: <i>p</i> value
	Young adults ^a	Older adults	
Age (years)	24.2 (3.1)	73.7 (6.2)	
Education (years)	15.5 (1.6)	16.8 (2.2)	<i>p</i> < .002
Modified Mini-Mental State Exam score (Mayeux, Stern, Rosen, & Leventhal, 1981)	54.6 (2.1)	54.2 (1.9)	n.s.
Verbal IQ	118.7 (14.1)	117.2 (7.6)	n.s.
Performance IQ	107.9 (14.9)	114.7 (10.8)	<i>p</i> < .02
Digits Forward	7.2 (1.3)	6.8 (1.3)	n.s.
Digits Backward	5.1 (1.3)	5.2 (1.2)	n.s.
Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961)	5.7 (5.2)	3.9 (3.4)	n.s.
SHORT CARE score (Gurland, Golden, Teresi, & Challop, 1984)			
Dementia (cutoff = 6)	NA	0.3 (0.5)	—
Depression (cutoff = 7)	NA	2.2 (1.4)	—

Note: Standard deviations are given in parentheses. NA = not applicable.

^aDemographic data are missing for 2 participants in this group.

At the completion of the first phase, the Electro-Cap (used for EEG recording; see the next section) was removed, and participants washed their hair. Approximately 10 min later, participants were retested on 20 high-confidence errors (questions responded to incorrectly and given a confidence rating of 5, 6, or 7), and 20 low-confidence errors (questions responded to incorrectly and given a confidence rating of 1, 2, or 3), as well as 20 questions for which responses had been omitted (data for the latter responses were not analyzed).

EEG recording

EEG was recorded from 62 scalp sites in accord with the extended 10-20 system (Sharbrough *et al.*, 1990). An Electro-Cap (Neuromedical Supplies; Compumedics USA, Inc., Charlotte, NC; with embedded sintered Ag/AgCl

electrodes and an averaged-mastoid reference) was used. Horizontal and vertical electrooculograms (EOGs) were recorded bipolarly with electrodes placed, respectively, at the outer canthi of both eyes and above and below the left eye. EOG and EEG were recorded continuously (Syn-amp amplifiers; Compumedics USA, Inc., Charlotte, NC; direct current, 100-Hz low-pass filter, 500-Hz digitization rate). Eye movement artifacts were corrected off-line (Semlitsch, Anderer, Schuster, & Presslich, 1986), and remaining artifacts were rejected manually. Impedances were kept below 5k Ω .

Results

In all analysis of variance comparisons (both behavioral and ERP data), we computed η_p^2 as a measure of effect size. For comparisons evaluated via *t* tests, Cohen's *d* was

Table 2. Sequence for a Single Trial in the Initial Semantic-Task Phase

1. "Please Wait" display (experimenter starts trial)
2. Prequestion interval (800- to 1,200-ms variable delay)
3. Question (participant has unlimited time to respond; experimenter records oral response by typing and marks its correctness)
4. Confidence screen (omitted following an "I don't know" response; participant has unlimited time to enter a rating from 1 to 7 using a keyboard)
5. Delay (500 ms)
6. Tone (300 ms) indicating that feedback will be displayed
7. Delay (1,500 ms)
8. Feedback (500 ms; the correct answer is displayed on the computer screen in green if the participant's response was correct and in red if it was incorrect)
9. Familiarity screen (omitted following a correct response; participant has unlimited time to indicate whether he or she is familiar with the correct response—1 = *familiar*, 2 = *unfamiliar*)

Note: Event-related potential recordings began 200 ms prior to the feedback. The trial sequence in the surprise retest phase was the same as shown here, except that the confidence and familiarity screens were not included.

calculated (for procedures, spreadsheets, and SPSS scripts to compute η_p^2 and Cohen's d , see Lakens, 2014). For all mean values reported, the 95% confidence intervals (CIs) are indicated.

Behavioral data

Semantic memory performance on the initial test.

The older adults had better semantic memory for the general-knowledge questions than did the young adults. The older participants' mean proportion correct was .41 ($SE = .016$, 95% CI = [.38, .44]), whereas the young participants' mean proportion correct was .26 ($SE = .01$, 95% CI = [.23, .28]), $t(87) = 7.28$, $d = 1.54$, $p < .0001$.

Overall, the older adults were more confident in their answers than the young adults (older adults: $M = 4.66$, $SE = 0.07$, 95% CI = [4.53, 4.80]; young adults: $M = 4.38$, $SE = 0.07$, 95% CI = [4.25, 4.51]; scores on a 7-point scale), $F(1, 87) = 8.67$, $MSE = 0.39$, $\eta_p^2 = .09$, $p = .004$. A similar result was obtained by Eich et al. (2013), although Cyr and Anderson (2013) reported the opposite finding. As expected, confidence was higher for items answered correctly ($M = 5.81$, $SE = 0.04$, 95% CI = [5.72, 5.89]) than for those answered incorrectly ($M = 3.23$, $SE = 0.07$, 95% CI = [3.10, 3.36]), $F(1, 87) = 1,812.22$, $MSE = 0.16$, $\eta_p^2 = .95$, $p < .0001$. For the confidence ratings, there was no interaction between correctness and age group ($F < 1$).

Error correction. All analyses reported in this section were conducted on items that had initially been answered incorrectly and for which corrective feedback had been given. The dependent variable was recall of the corrective feedback on the final test. Figure 1 shows that there was a main effect of age group on recall, such that older adults corrected more errors (proportion of total errors corrected: $M = .73$, $SE = .02$, 95% CI = [.69, .77]) than did the young adults ($M = .66$, $SE = .02$, 95% CI = [.62, .69]), $F(1, 87) = 6.33$, $MSE = 0.02$, $\eta_p^2 = .07$, $p < .014$. This finding is of considerable interest because error correction is new learning, and the older participants exhibited better, not worse, new learning than the young participants.

There was a main effect of confidence in the original error on later, postfeedback, recall of the correct answer, $F(1, 87) = 142.00$, $MSE = 0.01$, $\eta_p^2 = .62$, $p < .0001$; high-confidence errors were corrected with a higher probability ($M = .79$, $SE = .01$, 95% CI = [.76, .82]) than were low-confidence errors ($M = .59$, $SE = .02$, 95% CI = [.55, .63]). This is the hypercorrection effect. Both groups showed a reliable hypercorrection effect—young adults: $t(43) = 9.41$, $d = 1.42$, $p < .0001$; older adults: $t(44) = 7.30$, $d = 1.09$, $p < .0001$. Critically, as shown in Figure 1, there was an interaction between age group and confidence, $F(1, 87) = 5.76$, $MSE = 0.012$, $\eta_p^2 = .06$, $p = .02$, indicating that the difference in final correct recall between high- and

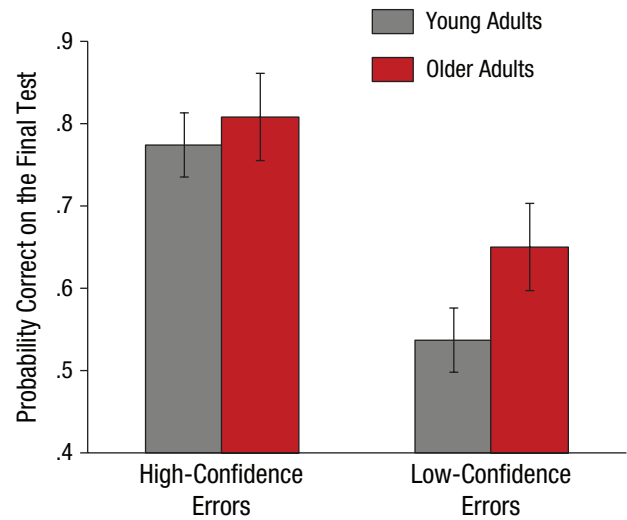


Fig. 1. Probability correct on the final surprise test for items that were high- and low-confidence errors on the initial semantic task, separately for the two age groups. Error bars indicate 95% confidence intervals. The data have been averaged across participants in each age group.

low-confidence errors was greater for young participants than older participants. That is, young, relative to older, participants hypercorrected to a greater extent.

To understand the interaction, we performed planned-comparison t tests. Importantly, while the proportion of high-confidence errors that were corrected was the same for older adults ($M = .81$, $SE = .02$, 95% CI = [.77, .85]) and young adults ($M = .77$, $SE = .02$, 95% CI = [.74, .81]), $t = 1.22$, the proportion of low-confidence errors that were corrected was greater for the older participants ($M = .65$, $SE = .03$, 95% CI = [.59, .70]) than for the young participants ($M = .54$, $SE = .02$, 95% CI = [.48, .59]), $t(87) = 3.02$, $d = 0.64$, $p = .003$. In addition, the difference between the proportions of high- and low-confidence errors that were corrected was, as predicted, reliably smaller for the older adults ($M = .16$, $SE = .02$, 95% CI = [.12, .19]) than for the young adults ($M = .24$, $SE = .03$, 95% CI = [.19, .28]), $t(87) = 2.40$, $d = 0.51$, $p = .019$.

In summary, the older adults performed better on the original general-information test than did the young adults and also corrected more of their errors on the surprise retest. Although older and young adults corrected their high-confidence errors to the same extent, the older adults corrected a greater proportion of their low-confidence errors, which resulted in a smaller hypercorrection effect for older than young adults.

ERP data

The ERPs we focus on in this section were collected while feedback to the errors was being presented. ERPs were recorded with a prestimulus baseline of 200 ms and

a poststimulus epoch of 900 ms. All ERP-average conditions (i.e., high-confidence, correct; high-confidence, incorrect; low-confidence, correct; low-confidence, incorrect) had at least five trials per participant, and were analyzed using the averaged amplitude of the P3a component. Here we report the ERPs to corrective feedback following errors. See the Supplemental Material available online for information on the ERPs to feedback following correct answers given high- or low-confidence judgments.

Because older adults typically exhibit P3 latencies that are prolonged relative to those of younger adults (Goodin, Squires, Henderson, & Starr, 1978), we used separate measurement windows to compute the average voltages for the young (375–425 ms) and older (425–475 ms) adults. Although these windows are longer than those associated with the auditory P3a, rare visual events elicit P3a components with longer latency than those observed in the auditory modality (Cycowicz & Friedman, 2007). We chose to focus on ERPs at FCz, where Butterfield and Mangels (2003) observed the maximal P3a in their study. Note that, in line with Butterfield and Mangels's (2003) data, Figure 2 (middle column) shows that a fronto-central distribution of the P3a was observed in both young and older adults.

Confidence had a main effect on the amplitude of the P3a component, $F(1, 87) = 162.51$, $MSE = 5.40$, $\eta_p^2 = .65$, $p < .0001$, as was shown originally by Butterfield and Mangels (2003). Relative to low-confidence-error corrective feedback ($M = 12.27 \mu\text{V}$, $SE = 0.71$, 95% CI = [10.88, 13.71]), high-confidence-error feedback ($M = 16.70 \mu\text{V}$, $SE = 0.73$, 95% CI = [15.31, 18.24]) produced greater P3a magnitudes.

Most important, as depicted in Figure 2, age group interacted with confidence, $F(1, 87) = 7.86$, $MSE = 5.40$, $\eta_p^2 = .08$, $p = .006$. To deconstruct the interaction, we performed planned-comparison t tests. Both groups showed a reliable difference between P3a magnitude to high- and low-confidence-error feedback—young adults: $M = 5.42 \mu\text{V}$, $SE = 0.52$, 95% CI = [4.51, 6.50], $t(43) = 10.43$, $d = 1.57$, $p < .0001$; older adults: $M = 3.47 \mu\text{V}$, $SE = 0.47$, 95% CI = [2.54, 4.35], $t(44) = 7.44$, $d = 1.11$, $p < .0001$. The pattern of results was consistent with the behavioral data: The young adults ($M = 15.54 \mu\text{V}$, $SE = 0.86$, 95% CI = [13.85, 17.11]) and the older adults ($M = 17.84 \mu\text{V}$, $SE = 1.15$, 95% CI = [15.60, 20.18]) did not differ in P3a magnitude to high-confidence-error feedback, $t = 1.60$, $p = .11$, $d = 0.34$. By contrast, older adults ($M = 14.37 \mu\text{V}$, $SE = 1.10$, 95% CI = [12.34, 16.41]), relative to young adults ($M = 10.12 \mu\text{V}$, $SE = 0.84$, 95% CI = [8.57, 11.73]), produced a larger P3a to low-confidence-error feedback, $t(87) = 3.14$, $d = 0.67$, $p = .002$. In highly similar fashion to the behavioral data, the ERP data revealed that the difference in P3a magnitude between high- and low-confidence-error feedback was smaller in the older than the young adults, $t(87) = 2.80$, $d = 0.59$, $p = .006$.

To summarize, in both age groups, having high confidence, rather than low confidence, in an error led to greater P3a amplitude when feedback provided the correct response. This suggests that the young and older adults both paid attention to corrective feedback on the rare occasions when they had expressed strong belief in the correctness of their responses but were wrong. By contrast, the fact that young, relative to older, adults demonstrated a larger P3a difference between the feedback to high- versus low-confidence errors may indicate that the young adults paid little attention to the corrective feedback given to their low-confidence errors. Then again, the older adults directed their attention to an almost equivalent extent to the corrective feedback that was provided following high- and low-confidence errors.

Additional analyses

Previous research (Metcalf & Finn, 2011; and see Sitzman et al., 2015) has shown that when young adults were asked to make a second guess without having received feedback but after they had given a wrong answer, that guess was more likely to be correct for high- than for low-confidence errors. Presumably, they made use of their preexisting semantic knowledge in making their second guesses. A similar recruitment of preexperimental knowledge could have also been at play in the present experiment and could, potentially, account for the pattern of behavioral data. To explore this possibility, we looked at ratings that participants gave about their familiarity with the correct responses immediately after corrective feedback was provided. It should be noted that these familiarity ratings—unlike the data provided by Metcalfe and Finn (2011, 2012)—could be due to mere hindsight bias rather than indicating that people really had prefeedback knowledge, insofar as the present ratings were obtained only after the correct answer was provided. Furthermore, the familiarity ratings might have been biased by the fluency in processing the feedback as a result of the encoding processing that had occurred just a moment earlier: The ERP data indicated that the processing of the correct responses differed between conditions and across participant age groups. Caution, then, is needed in interpreting these data.

Even so, we computed the proportion of trials, in each confidence condition, in which participants said they were familiar with the feedback. These familiarity ratings mirrored the pattern seen in the behavioral data shown in Figure 1. In particular, the ratings were high for the feedback to high-confidence errors (and did not differ between groups: $M = 0.40$, $SE = 0.05$, 95% CI = [0.31, 0.50], for the young and $M = 0.39$, $SE = 0.05$, 95% CI = [0.30, 0.50], for the older adults), and were lower for the low-confidence errors, $F(1, 87) = 7.71$, $MSE = 0.05$, $\eta_p^2 = .08$, $p = .007$. Although the interaction between age group

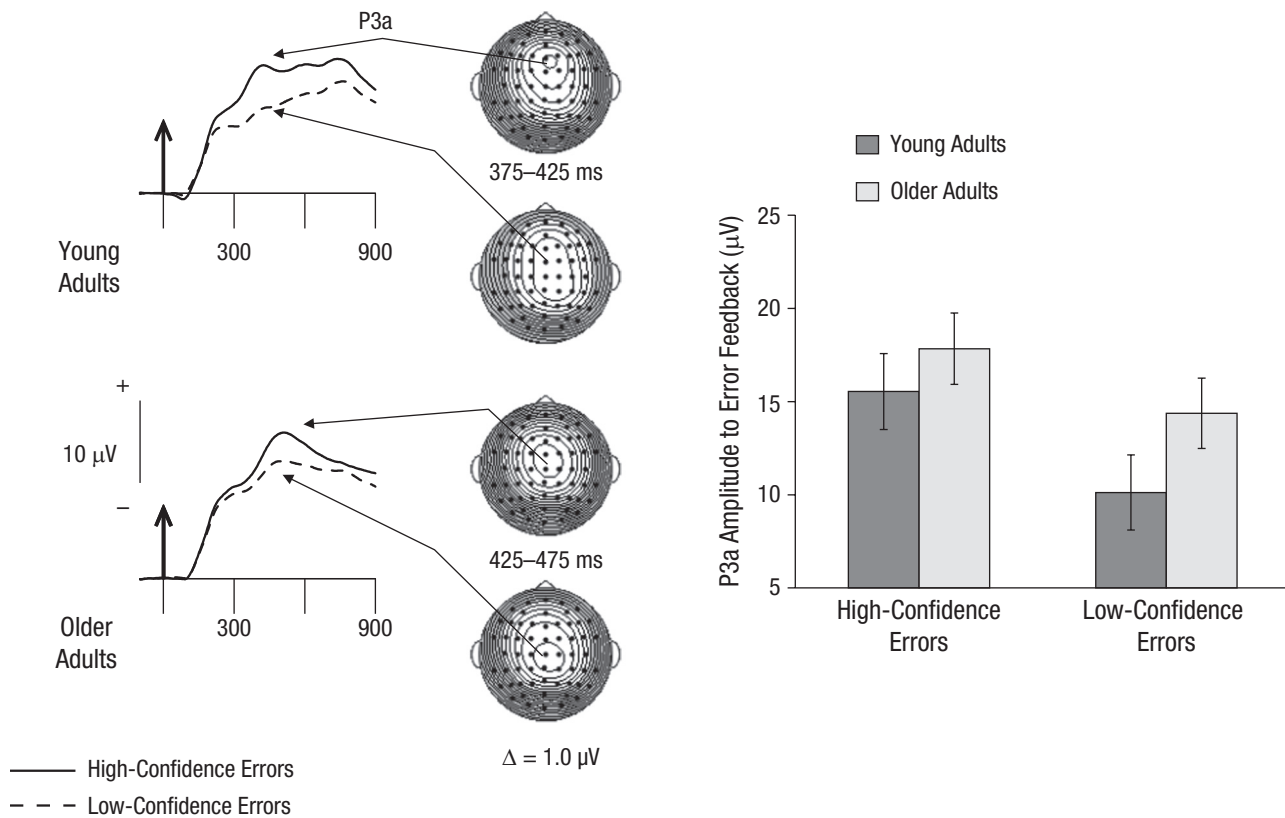


Fig. 2. Grand-mean event-related potentials (ERPs), scalp distributions, and averaged voltages associated with high- and low-confidence errors. The left column depicts the grand-mean ERPs at FCz elicited by the corrective feedback to high- and low-confidence errors, averaged across participants within each age group (young adults at the top, older adults at the bottom). The middle column illustrates the scalp distribution of P3a averaged voltages associated with the waveforms in the first column. The delta (Δ) of 1 indicates that the isopotential lines in the maps are separated by 1 μ V. Positive areas are unshaded, whereas negative areas are shaded, and the dots indicate the electrode locations. The right column presents the grand-mean P3a averaged voltages (based on the waveforms in the left column) to corrective feedback following high- and low-confidence errors, computed across participants in each age group. Error bars indicate 95% confidence intervals.

and confidence was not significant, $F(1, 87) = 1.807$, $\eta_p^2 = .020$, $p = .182$, the familiarity ratings for the low-confidence-error feedback were higher numerically for the older adults ($M = 0.35$, $SE = 0.02$, 95% CI = [0.31, 0.39]) than for the young adults ($M = 0.26$, $SE = 0.02$, 95% CI = [0.22, 0.30]). Despite the nonsignificant interaction, we contrasted the group differences separately by confidence level, finding that the between-group difference in familiarity was not significant for the feedback to high-confidence errors ($t < 1$), but was for the feedback to low-confidence errors, $t(87) = 2.70$, $d = 0.57$, $p = .008$. The similarity between the pattern of familiarity ratings and the pattern of later recall is striking and led us to wonder: Could familiarity account for the behavioral interaction seen earlier? If the familiarity ratings truly measured preexisting knowledge, and if the older adults, in particular, relied on that preexisting knowledge to make their final recall responses, this might affect our conclusions that the older adults learned better than the young adults in this paradigm.

To investigate the extent to which participants relied on familiarity, we separated the items into four classes: high-confidence-error feedback deemed familiar and unfamiliar, and low-confidence-error feedback deemed familiar and unfamiliar. We then computed the proportion of final recall for each category. As might be expected, familiarity with the correct response was associated with increased recall. However, the effect was not large. The mean difference in recall between familiar and unfamiliar items was .17 ($SE = .03$, 95% CI = [.13, .21]). It might be expected that this difference would be larger for the older than for the young adults if the older adults had relied more on their crystallized preexperimental knowledge, rather than on new learning. However, the difference was smaller for the older adults ($M = .09$, $SE = .03$, 95% CI = [.04, .14]) than for the young adults ($M = .27$, $SE = .03$, 95% CI = [.22, .32]), $t(87) = 4.96$, $d = 1.05$, $p < .0001$.

Finally, if the older adults had been relying on preexperimental knowledge, rather than new learning, they should have done particularly poorly on the unfamiliar

items, because they would have been unable to depend on their past knowledge store for these items. However, as shown in Figure 3, although recall on the familiar items was equal for the two groups, the older adults recalled the unfamiliar items particularly well ($M = .68$, $SE = .02$, 95% CI = [.62, .72]), and significantly better than did the young adults ($M = .56$, $SE = .03$, 95% CI = [.51, .59]), $t(87) = 3.29$, $d = 0.70$, $p < .001$.

The claim that prior knowledge has an impact on error correction deserves careful consideration. But although we do not deny the importance of a role for prior knowledge in error correction, the additional analyses presented here run counter to the possibility that the older adults, in this experiment, were merely relying on their prior knowledge, or were even relying on it more than the young adults: They were relying on it less. And, in the cases that most demanded new learning—the correct answers that participants claimed were unfamiliar to them—the older adults learned better than the young adults.

Discussion and Conclusion

These results indicate not only that older adults performed better on a general-information task that tapped into their factual knowledge, but also that they corrected their errors better than did young adults. They did not hypercorrect as much as did the young adults, but the smaller difference between the correction probability of high- versus low-confidence errors was not due to a processing deficit. Instead, our data indicated enhanced processing of the low-confidence errors: Older adults tended to correct all of their errors rather than just focusing on high-confidence errors.

The ERPs provide a window into the brain processing concurrent with the presentation of the feedback. Whereas the young adults evinced a large difference between the attention-related P3a (Friedman et al., 2001) to high- and low-confidence errors, in association with a large hypercorrection effect, this difference for the older adults, both in the P3a and in hypercorrection, was smaller. Older adults rallied their attention more effectively to all errors and, hence, learned better than younger adults.

A question that comes to mind is why, in this paradigm, older adults were able to overwrite their old response patterns and learn new information better than young adults, when it is widely accepted that new learning is particularly problematic for older adults. It is, of course, possible that the fact that this paradigm focuses on semantic rather than episodic memory (Tulving, 2002) is central. Perhaps the goodness of processing depends on the system engaged, and semantic encoding as well as storage is spared with aging (Mitchell, 1989).

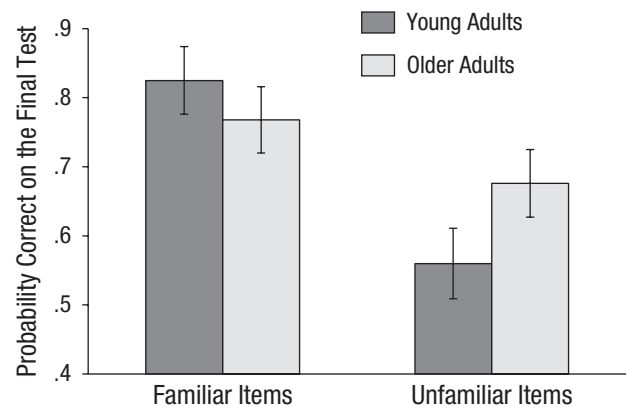


Fig. 3. Probability correct on the final surprise test for feedback items that were rated familiar and unfamiliar after feedback was provided. Error bars indicate 95% confidence intervals. The data have been averaged across participants in each age group.

There is another possibility, however. It is possible that the paradigm we used is special, in a laboratory context at least, because the new knowledge that participants are asked to learn is the truth, rather than arbitrary information. Older adults may be particularly motivated to learn the truth, and capable of engaging their attention to this end. In short, they may be highly epistemically motivated. Metcalfe (2015) has recently shown that people process information differently when they are asked to remember the true answers to factual questions compared with when they are asked to remember false answers. Nearly all experiments that have shown that older adults have more difficulty updating their memories than young adults have used to-be-learned materials that either were epistemically neutral or were patently wrong. For example, in Ruch's (1934) classic study illustrating a supposed learning deficit, people were asked to learn new, but wrong, answers to multiplication problems. Older adults exhibited considerable difficulty with this task: Younger adults remembered better than older adults "facts" such as $3 \times 4 = 2$. Older adults' poor performance may have stemmed not from a difficulty in learning but from a reluctance to rally their precious attentional resources in the service of false or useless information. Older adults also exhibited difficulty learning deviant variations of well-known fairy tales (Attali & Dalla Barba, 2012), perhaps because they thought the variations were simply wrong. And they have difficulty learning arbitrary word pairs in which no truth value is enlisted.

In support of the possibility that truth and relevance matter to older adults, Castel (2005) showed that although older adults performed poorly when asked to learn object-price pairs (of grocery items) that were unrealistic, their memory performance equaled that of young adults when the object-price pairings were realistic. Further research is needed to investigate this possibility, but we

suggest that older adults are capable of rallying their attentional resources as well as, and sometimes better than, young adults. But they do so selectively. One factor that may be central is that the corrective to-be-learned information be factually correct. Older adults may be unwilling or unable to recruit their efforts to learn irrelevant mumbo jumbo, but, as the present study demonstrates, they can and will engage their attention and effort to learn the truth.

Author Contributions

J. Metcalfe, D. Friedman, and A. Radin conceived the experiment, and L. Casal-Roscum performed the statistical analyses. J. Metcalfe and D. Friedman wrote the manuscript. A. Radin and L. Casal-Roscum also ran the experiment.

Acknowledgments

We thank Judy Xu for her help.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by National Institute on Aging Grant AG005213 (to D. Friedman) and James S. McDonnell Grant 220020166 (to J. Metcalfe).

Supplemental Material

Additional supporting information can be found at <http://pss.sagepub.com/content/by/supplemental-data>

Note

1. In Cyr and Anderson's (2013) Experiment 1, the young adults outperformed the older adults on the retest. In this study, the entire pool of 150 items was retested regardless of response correctness, and the order of presentation was identical during the initial test (study) and the final test, a procedural feature that may have provided contextual information that helped the young more than the older participants. Order of presentation in the retest was not specified for the experiment by Sitzman et al. (2015). In it, older adults performed better than young adults overall and also in correcting low-confidence errors. Both studies showed smaller hypercorrection effects for older than for young adults.

References

- Attali, E., & Dalla Barba, G. (2012). Confabulation in healthy aging is related to poor encoding and retrieval of over-learned information. *Aging, Neuropsychology, and Cognition, 20*, 339–355.
- Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–409). New York, NY: Oxford University Press.
- Baudena, P., Halgren, E., Heit, G., & Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalography & Clinical Neurophysiology, 94*, 251–264.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.
- Botwinick, J. (1984). *Aging and behavior: A comprehensive integration of research findings*. New York, NY: Springer.
- Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review, 18*, 1238–1244.
- Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research, 17*, 793–817.
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491–1494.
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1*, 69–84.
- Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20*, 718–721. doi:10.1037/0882-7974.20.4.718
- Cycowicz, Y. M., & Friedman, D. (2007). Visual novel stimuli in an ERP novelty oddball paradigm: Effects of familiarity on repetition and recognition memory. *Psychophysiology, 44*, 11–29.
- Cyr, A. A., & Anderson, N. D. (2013). Updating misconceptions: Effects of age and confidence. *Psychonomic Bulletin & Review, 20*, 574–580.
- Eich, T. S., Stern, Y., & Metcalfe, J. (2013). The hypercorrection effect in younger and older adults. *Aging, Neuropsychology, and Cognition, 20*, 511–521. doi:10.1080/13825585.2012.754399
- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 224–240.
- Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review, 16*, 88–92.
- Fazio, L. K., & Marsh, E. J. (2010). Correcting false memories. *Psychological Science in the Public Interest, 21*, 801–803.
- Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews, 25*, 355–373.
- Friedman, D., Nessler, D., Kulik, J., & Hamberger, M. (2011). The brain's orienting response (novelty P3) in patients with unilateral temporal lobe resections. *Neuropsychologia, 49*, 3474–3483. doi:10.1016/j.neuropsychologia.2011.08.023
- Goodin, D. C., Squires, K. C., Henderson, B. H., & Starr, A. (1978). Age-related variations in evoked potentials to auditory stimuli in normal human subjects. *Electroencephalography & Clinical Neurophysiology, 44*, 447–458.
- Gurland, B., Golden, R. R., Teresi, J. A., & Challop, J. (1984). The SHORT-CARE: An efficient instrument for the

- assessment of depression, dementia and disability. *Journal of Gerontology*, 39, 166–169.
- Iwaki, N., Matsushima, H., & Kodaira, K. (2013). Hypercorrection of high confidence errors in lexical representations. *Perceptual & Motor Skills*, 117, 219–235.
- Jacoby, L. L., Hessels, S., & Bopp, K. L. (2001). Proactive and retroactive effects in memory performance: Dissociating recollection and accessibility bias. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 35–54). Washington, DC: American Psychological Association.
- Knight, R. (1984). Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography & Clinical Neurophysiology*, 59, 9–20.
- Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology*, 68, 522–528.
- Lakens, D. (2014). *Calculating confidence intervals for Cohen's d and eta-squared using SPSS, R, and Stata*. Retrieved from <http://daniellakens.blogspot.ca/2014/06/calculating-confidence-intervals-for.html>
- Mayeux, R., Stern, Y., Rosen, J., & Leventhal, J. (1981). Depression, intellectual impairment, and Parkinson's disease. *Neurology*, 31, 645–650.
- McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology*, 41, 175–192.
- Metcalf, J. (2015). *Correcting errors with true feedback versus overwriting correct answers with errors*. Manuscript submitted for publication.
- Metcalf, J., Butterfield, B., Habeck, C., & Stern, Y. (2012). Neural correlates of people's hypercorrection of their false beliefs. *Journal of Cognitive Neuroscience*, 24, 1571–1583.
- Metcalf, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 437–448.
- Metcalf, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction*, 22, 253–261.
- Metcalf, J., & Miele, D. B. (2014). Hypercorrection of high confidence errors: Prior testing both enhances delayed performance and blocks the return of the errors. *Journal of Applied Research in Memory and Cognition*, 3, 189–197.
- Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 31–49.
- Nelson, T., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, 19, 338–368.
- Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography & Clinical Neurophysiology*, 67, 360–371.
- Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology*, 14, 330–345.
- Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, 4, 193–202.
- Ruch, F. L. (1934). The differentiative effects of age upon human learning. *Journal of General Psychology*, 11, 261–286.
- Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, 23, 695–703.
- Sharbrough, F., Chatrian, G. E., Lesser, R. P., Luders, H., Nuwer, M., & Picton, T. W. (1990). *Guidelines for standard electrode position nomenclature*. Bloomfield, CT: American EEG Society.
- Sitzman, D. M., Rhodes, M. G., Tauber, S. K., & Licalde, V. R. (2015). The role of prior knowledge in error correction for younger and older adults. *Aging, Neuropsychology, and Cognition*, 22, 502–516. doi:10.1080/13825585.2014.993302
- Staudinger, U. M., Cornelius, S. W., & Baltes, P. B. (1989). The aging of intelligence: Potentials and limits. *Annals of the American Academy of Political and Social Science*, 503, 43–59.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25.
- Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, 9, 408–426.