Published in final edited form as:

Psychol Sci. 2015 December ; 26(12): 1833–1842. doi:10.1177/0956797615597912.

On Teaching Old Dogs New Tricks

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

¹Columbia University Department of Psychology

²Division of Cognitive Neuroscience, New York State Psychiatric Institute

Abstract

Although older adults rarely outperform young adults on learning tasks, here they surpassed their younger counterparts both on answering more semantic-memory general-information questions correctly, but also on 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 (ERPs) 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 P3a’s 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 error and regardless of their familiarity with the answers.

INTRODUCTION

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 they, by virtue of having lived longer, 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, investigated here, that not only are older healthy adults better than young adults at answering general information questions in the first place, but, when they do make a mistake the older adults might also 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, it directly contravenes 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 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, until people 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 to obtain the same number of incorrect responses, that is, their 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, while 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. 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 older adults had to be asked more questions than 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 Footnote 1). Second, the smaller hypercorrection effect among the older adults could have come about in two ways: (1) because the high confidence errors were corrected *less* (suggesting a deficit) or (2) because the low-confidence errors were corrected *more* (suggesting enhanced memory).

To further investigate these differences, we conducted an 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;

¹In Cyr and Anderson’s (2013) Experiment 1 the entire pool of 150 items was retested regardless of original correctness. But, the order of test presentation was identical to the study presentation order, a procedural feature that may have provided contextual information that helped the young more than the older participants, resulting in better memory for the former than the latter. Order of presentation in the retest was not specified in Sitzman et al., (2015), but older adults performed better than the young including correcting low confidence errors better.

Fazio & Marsh, 2010; c.f., Metcalfe et al., 2012; Metcalfe & Finn, 2012). 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 to those that are forgotten (e.g., Fabiani & Donchin, 1995; Paller, Kutas, & Mayes, 1985; Ranganath & Rainer, 2003). Further, the hippocampus, a critical structure in memory, is also thought to be a component 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, an fMRI study (Metcalf et al., 2012) showed greater anterior cingulate cortex (ACC) activation associated with the feedback to high- as compared 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. 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 more likely to miss detecting the tone during the presentation of corrective feedback to high than 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 difference in hypercorrection effect could also have resulted not because the older adults paid *less* attention to the high-confidence errors, but, rather because they had superior processing on the 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 high-confidence errors and perhaps also an attentional P3a to low-confidence errors.

Our hypotheses, then, were that (1) older adults would have greater semantic knowledge on the pretest than young adults, (2) 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 than young adults—showing superior new learning, (3) replicating past research, older adults would show a smaller hypercorrection effect than young adults, (4) the smaller hypercorrection effect would be due to better memory for the low-confidence error feedback rather than worse memory for the high-confidence error feedback, and (5) there would not be a decrease in the high-confidence error P3a in older adults, and there might be enhanced attentional processing (reflected by the P3a) for the low-confidence-error feedback for these participants. The alternative ‘deficit’ hypothesis indicates that although semantic memory should be better for older adults on the pretest, error correction should be worse. It predicts a

smaller hypercorrection effect for older adults attributable to older adults failing to pay attention to the feedback to high-confidence errors, which would then be reflected in little or no P3a to the feedback to high-confidence errors.

METHOD

Participants

Forty-four young adults (25 women) 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) 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 (see Table 1). Ten additional young and 10 older adults were eliminated due to 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-group 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). These investigators recruited 25 young adults in their first experiment, and ended up with 20 with usable data; they recruited 23 participants and ended up with 20 with usable data in their second experiment. These experiments demonstrated a confidence-related effect on the P3a. Hence, we aimed for approximately 40–50 per Age Group to allow evaluation of a potential group interaction related to confidence and the P3a.

The experiment typically lasted 4 hours because it took a long time to accumulate a sufficient number of high-confidence errors (we required at least 20), as they are inherently rare. All participants were native English speakers, with normal or corrected-to-normal vision, no history of neurological or psychiatric disorders, and free from medications known to affect the central nervous system.

Older adults were pre screened with an extensive telephone inventory which was then 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 normally aging. All participants signed informed consent according to the criteria of the New York State Psychiatric Institute's Institutional Review Board and were paid \$15/hr.

Stimuli and Procedures

The stimuli were 439 general-information questions from 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 of 3 to 14 letters in length (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. EEG was recorded only during the first phase. General information questions were presented in the center of the computer

screen. Participants were given an unlimited amount of time to respond verbally to each question. Participants were encouraged to guess if they were not sure for all questions, but were allowed to say “I don’t know,” that is, to omit responses. The experimenter recorded the participants’ responses by typing their answer on a keyboard. For all responses except omits, participants rated their confidence in the correctness of their response on a scale ranging from 1 (least confident) to 7 (most confident) on a keyboard. Participants were encouraged to use the entire scale.

Immediately following the confidence rating or omit response, a central fixation point appeared. Then, a tone was delivered via speaker to announce the onset of the visual feedback, which then provided the correct answer. If the participant’s response had been an error, s/he then entered their familiarity with the correct answer.

There were 44 questions in each block, with a short break, of approximately 5 min, after each block. 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). There was no difference between young- and older-adult participants in this number ($t < 1$).

At completion of the first phase, the Electrocap (see EEG methods below) was removed and participants washed their hair. Approximately 10 minutes later, participants were retested on 20 high-confidence errors (questions to which the participant responded incorrectly, but rated their confidence as 5–7), and 20 low confidence errors (rated 1–3), as well as 20 omits that were not analyzed.

Electroencephalographic (EEG) Recording

EEG was recorded from 62 scalp sites (sintered Ag/AgCl) in accord with the extended 10–20 system (Sharbrough et al., 1990) using an Electrocap (Neuromedical Supplies) and an averaged-mastoid reference. Horizontal and vertical electrooculogram (EOG) 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 (Synamp amplifiers; DC; 100-Hz low pass filter; 500-Hz digitization rate). Eye movement artifacts were corrected offline (Semlitsch, Anderer, Schuster, & Presslich, 1986) and remaining artifacts were rejected manually. Impedances were kept below 5k Ω .

RESULTS

Behavioral data

In all comparisons (both behavioral and ERP), η^2 , a measure of effect size (available in SPSS) was computed. For comparisons evaluated via t tests, Cohen’s d was calculated (for procedures, spreadsheets and SPSS scripts to compute η^2 , and Cohen’s d see: <http://daniellakens.blogspot.ca/2014/06/calculating-confidence-intervals-for.html>). For all mean values reported in the text and figures, the 95% CI’s were calculated and listed in the results or depicted in the figures.

Semantic Memory Performance on The Initial Test—Older adults had better semantic memory for the general knowledge questions than did the young participants. The older participants' mean proportion correct was 0.41 (SE=0.016; 95% CI [0.38, 0.44]), whereas the mean proportion correct for the young participants was 0.26 (SE=0.01; 95% CI [0.23, 0.28]), $t(87) = 7.28$, $d = 1.54$, $P < 0.0001$.

Older adults were more confident overall than young adults (respectively, 4.66, SE=0.07; 95% CI [4.53, 4.80] vs. 4.38, SE=0.07; 95% CI [4.25, 4.51] on a 7 point scale), $F(1,87) = 8.67$, MSE = 0.39, $\eta^2 = 0.09$, $P = 0.004$. A similar result was found by Eich et al., (2013), although the opposite finding was reported by Cyr & Anderson, (2013). As expected, confidence was higher for items on which people had given the correct answer ($M = 5.81$, SE=0.04; 95% CI [5.72, 5.89]) than those on which they had made an error ($M = 3.23$, SE = 0.07; 95% CI [3.10, 3.36]), $F(1,87) = 1812.22$, MSE = 0.16, $\eta^2 = 0.95$, $P < 0.0001$. For the confidence ratings, there was no interaction between correctness and group ($F < 1$).

Error Correction—All analyses reported in this section were conducted on items that had initially been answered incorrectly and had been given corrective feedback. The dependent variable in this section is recall of the corrective feedback on the final test.

Figure 1 shows that there was a main effect of Age Group, such that older participants corrected more errors (proportion of total errors corrected: $M = 0.73$, SE = 0.02, 95% CI [0.69, 0.77]) than did younger participants, ($M = 0.66$, SE = 0.02; 95% CI [0.62, 0.69], ($F(1,87) = 6.33$, MSE = 0.02, $\eta^2 = 0.07$, 95% CI $P = 0.014$). This finding is of considerable interest because error correction is new learning, and the older participants corrected better, not worse than the young participants.

There was a main effect of confidence in the original error on later, post feedback, recall of the correct answer ($F(1,87) = 142.00$, MSE = 0.01, $\eta^2 = 0.62$, $P < 0.0001$) such that high-confidence errors were corrected with a higher probability ($M = 0.79$, SE = 0.01, 95% CI [0.76, 0.82]) than were low-confidence errors ($M = 0.59$, SE = 0.02, 95% CI [0.55, 0.63]). This is the hypercorrection effect. Both groups showed a reliable hypercorrection effect ($t(43)_{\text{young}} = 9.41$, $d = 1.42$, $P < 0.0001$; $t(44)_{\text{older}} = 7.30$, $d = 1.09$, $P < 0.0001$). Critically, as is shown in Figure 1, there was an interaction between Age Group and confidence ($F(1,87) = 5.76$, $\eta^2 = 0.06$, $P = 0.02$), indicating that the difference in final correct recall between high and 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, planned-comparison t -tests were performed. Importantly, while the proportion of *high* confidence errors that were corrected was the same for older ($M = 0.81$, SE = 0.02, 95% CI [0.77, 0.85]) and young ($M = 0.77$, SE = 0.02, 95% CI [0.74, 0.81]) participants, $t = 1.22$, the proportion of *low confidence* errors that were corrected was *greater* for the older ($M = 0.65$, SE = 0.03, 95% CI [0.59, 0.70]) than for the young participants ($M = 0.54$, SE = 0.02, 95% CI [0.48, 0.59]), $t(87) = 3.02$, $P = 0.003$. In addition, the difference between high- and low-confidence errors that were corrected was, as predicted, reliably smaller for the older ($M = 0.16$, SE = 0.02, 95% CI [0.12, 0.19]) than young adults ($M = 0.24$, SE = 0.03, 95% CI [0.19, 0.28]), $t(87) = 2.40$, $P = 0.019$.

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

ERP Data

The ERPs we focus on below were collected while the feedback on the errors was being presented. ERPs were recorded with a pre-stimulus baseline of 200 ms and a post-stimulus epoch of 900 ms. All ERP average conditions had at least 5 trials per participant, and were analyzed using the averaged amplitude of the P3a component. Here we report the ERPs to corrective feedback following errors. The ERPs to the feedback to correct answers, as a function of whether the person had expressed that correct answer with high or low confidence, are presented in the supplementary materials (Figure S1 and associated analyses).

Because older adults typically evince P3 latencies that are prolonged relative to those of younger adults (Goodin, Squires, Henderson, & Starr, 1978), separate measurement windows were used to compute the averaged voltages for the young (375–425 ms) and older (425–475 ms) adults. Although these windows appear longer than those associated with the auditory P3a, rare, visual events elicit P3a's with longer latency than those observed in the auditory modality (Cycowicz & Friedman, 2007). Electrode choice (FCz) was driven by where the P3a had been observed to be maximal in the data of Butterfield and Mangels, (2003).

P3a Elicited by Error Feedback at FCz—There was a main effect of confidence on the P3a component ($F(1,87) = 162.51$, $MSE = 5.40$, $\eta^2 = 0.65$, $P < 0.0001$), as had been shown originally by Butterfield and Mangels (2003). Relative to low-confidence error corrective feedback ($M = 12.27 \mu V$, $SE = 0.71$, 95% CI [10.88, 13.71]), high-confidence error feedback ($M = 16.70 \mu V$, $SE = 0.73$, 95% CI [15.31, 18.24]) produced greater P3a magnitudes.

Most important, as depicted in Figure 2, Age Group interacted with Confidence ($F(1,87) = 7.86$, $\eta^2 = 0.08$, $P = 0.006$). To deconstruct the interaction, planned-comparison t-tests were performed. Both groups showed a reliable difference between P3a magnitude to high- and low-confidence error feedback (young $M = 5.42 \mu V$, $SE = 0.52$; 95% CI [4.51, 6.50], $t(43) = 10.43$, $d = 1.57$, $P < 0.0001$; older $M = 3.47 \mu V$, $SE = 0.47$; 95% CI [2.54, 4.35], $t(44) = 7.44$, $d = 1.11$, $P < 0.0001$). In accord with the behavioral data, young ($M = 15.54 \mu V$, $SE = 0.86$, 95% CI [13.85, 17.11]) and older ($M = 17.84 \mu V$, $SE = 1.15$, 95% CI [15.60, 20.18]) adults did not differ in P3a magnitude to high-confidence error feedback ($t = 1.60$, $P = 0.11$, $d = 0.34$). By contrast, older ($M = 14.37 \mu V$, $SE = 1.10$, 95% CI [12.34, 16.41]) relative to young ($M = 10.12 \mu V$, $SE = 0.84$, 95% CI [8.57, 11.73]), adults produced a larger P3a to low-confidence error feedback ($t(87) = 3.14$, $d = 0.67$, $P = 0.002$). In highly similar fashion to the behavioral data, the P3a-magnitude difference between high- and low-confidence error feedback was *smaller* in the older than young adults ($t(87) = 2.80$, $d = 0.59$, $P = 0.006$).

To summarize, in both age groups, having high-confidence in an error led to greater amplitude P3a when feedback provided the correct response relative to when a low-confidence judgment resulted in an error. This suggests that both young and older adults paid attention to corrective feedback to the rare events in which 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 paid little attention to the corrective feedback given to their low-confidence errors. By contrast, older adults directed their attention to an almost equivalent extent to the corrective feedback provided by both 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 pre-existing semantic knowledge. A similar recruitment of pre-experimental 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 the 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 & Finn (2011; 2012)—could be due to mere hindsight bias rather than indicating that people really had pre-feedback knowledge, since the present ratings were obtained only *after* the correct answer was provided. Furthermore, the familiarity ratings might have reflected people's assessment of the fluency of the judged item that resulted from 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 groups. Caution, then, is needed in interpreting these data.

Even so, we computed the proportion of trials, in each 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 with a mean of 0.40 (SE=0.05, 95% CI [0.31, 0.50]) for the young adults and 0.39 (SE=0.05, 95% CI [0.30, 0.50]) for the older adults), and were lower for low confidence errors, $F(1,87) = 7.71$, $MSE = 0.05$, $\eta^2 = 0.08$, $P = 0.007$. Although the interaction between group and confidence was not significant, $F(1,87) = 1.807$, $\eta^2 = 0.020$, $P = 0.182$, the familiarity ratings on the low confidence errors were higher numerically for the older ($M = 0.35$, $SE = 0.02$, 95% CI [0.31, 0.39]) than for the young adults ($M = 0.26$, $SE = 0.02$, 95% CI [0.22, 0.30]). If, despite the non-significant interaction, one contrasts the group differences separately by confidence level, the between-group difference in familiarity is not significant for the high- ($t < 1$), but is for the low confidence errors, $t(87) = 2.70$, $d = 0.57$, $P = 0.008$. The similarity of the patterns of familiarity ratings to 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 really measured pre-existing knowledge and, if the older adults, in particular, relied on that pre-

existing knowledge to make their final recall responses, it might impact our conclusions that the older adults *learned* better than the young adults in this paradigm.

To investigate the extent to which participants were relying on familiarity, we separated the data of each participant into four classes: items that were high and low confidence errors and were deemed either familiar or unfamiliar, and computed proportion final recall in each. As might be expected, being familiar with the correct response was related to increased recall. However, the effect was not large (difference = 0.17, SE=0.03, 95% CI [0.13, 0.21] between familiar and unfamiliar). It might be expected that if the older adults were relying more on their crystallized pre-experimental knowledge, rather than on new learning, they would show a larger difference than would young adults. However, the difference was *smaller* for the older adults (0.09 between items deemed familiar and unfamiliar, SE=0.03, 95% CI [0.04, 0.14]) than for the young adults (0.27, SE=0.03, 95% CI [0.22, 0.32]), $t(87) = 4.96$, $d = 1.05$, $P < 0.0001$.

Finally, if the older adults were relying on pre-experimental knowledge, rather than new learning, they should have done particularly poorly on the unfamiliar items, since it was on these items that they would be unable to depend on that past knowledge store. However, as is shown in Figure 3, while recall was equal for the two groups on the familiar items, the older adults recalled the *unfamiliar* items particularly well ($M = 0.68$, SE=0.02, 95% CI [0.62, 0.72]), and significantly *better* than did the younger adults ($M = 0.56$, SE=0.03, 95% CI [0.51, 0.59]), $t(87) = 3.29$, $d = 0.70$, $P < 0.001$).

The claim that prior knowledge has an impact on error correction deserves careful consideration. But while not denying 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 just 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 responses that participants claimed were *unfamiliar* to them—the older adults did better than the young adults.

DISCUSSION AND CONCLUSION

These results indicate that older adults not only performed better on a general information task that tapped into their factual knowledge, but they also corrected their errors better than did young adults. They did not hypercorrect as much as did young adults but this 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 provided a window into the brain processing concurrent with the presentation of the feedback. While the young 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 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).

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 people were being asked to learn was *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 (under revision) has recently shown that people process differently when they are asked to remember the true answers to factual questions as compared to remembering 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 were either epistemically neutral or were patently wrong. For example, in the classic study by Ruch (1934) illustrating a so-called learning deficit, people were asked to learn new, but wrong, answers to multiplication problems. Older adults exhibited considerable difficulty with this: 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 because of 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 just wrong. And, they have difficulty with arbitrary paired associates, in the A–B A–C paradigm, 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 when older adults were asked to learn object price pairs (of grocery items) that were unrealistic they did badly, when the object-price pairings were realistic their memory performance equaled that of young adults. Further research is needed to investigate this possibility, but we suggest that older adults are capable of rallying their attentional resources as well 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. They 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING: Supported by NIA grant #AG005213 (DF) and McDonnell grant #220020166 (JM). We thank Judy Xu for her help.

References

- Attali E, Dalla Barba G. Confabulation in healthy aging is related to poor encoding and retrieval of over-learned information. *Aging, Neuropsychology, and Cognition*. 2012; 20:339–355.
- Balota, DA.; Dolan, PO.; Ducheck, JM. Memory changes in healthy older adults. In: Tulving, E.; Craik, FIM., editors. *The Oxford handbook of memory*. NY, NY: Oxford University; 2000. p. 395–409.
- Baudena P, Halgren E, Heit G, Clarke JM. Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalography and Clinical Neurophysiology*. 1995; 94(4):251–264. [PubMed: 7537197]
- Botwinick, J. *Aging and behavior: A comprehensive integration of research findings*. New York, NY: Springer; 1984.
- Butler AC, Fazio LK, Marsh EJ. The hypercorrection effect persists over a week, but high confidence errors return. *Psychonomic Bulletin and Review*. 2011; 18:1238–1244. [PubMed: 21989771]
- Butterfield B, Mangels JA. Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*. 2003; 17(3):793–817. [PubMed: 14561464]
- Butterfield B, Metcalfe J. Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2001; 27(6):1491–1494.
- Butterfield B, Metcalfe J. The correction of errors committed with high confidence. *Metacognition and Learning*. 2006; 1:69–84.
- Castel AD. Memory for grocery prices in younger and older adults: the role of schematic support. *Psychology and Aging*. 2005; 20(4):718–721.10.1037/0882-7974.20.4.718 [PubMed: 16420146]
- Cycowicz YM, Friedman D. Visual novel stimuli in an ERP novelty oddball paradigm: effects of familiarity on repetition and recognition memory. *Psychophysiology*. 2007; 44(1):11–29. [PubMed: 17241137]
- Cyr AA, Anderson ND. Updating misconceptions: Effects of age and confidence. *Psychonomic Bulletin and Review*. 2013; 20:574–580. [PubMed: 23254575]
- Eich TS, Stern Y, Metcalfe J. The hypercorrection effect in younger and older adults. *Neuropsychol Dev Cogn B Aging Neuropsychology and Cognition*. 2013; 20(5):511–521.10.1080/13825585.2012.754399
- Fabiani M, Donchin E. Encoding processes and memory organization: a model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1995; 21(1):224–240.
- Fazio LK, Marsh EJ. Correcting false memories. *Psychological Science in the Public Interest*. 2010; 21:801–803.
- Friedman D, Cycowicz YM, Gaeta H. The novelty P3: an event related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*. 2001; 25(4):355–373. [PubMed: 11445140]
- Friedman D, Nessler D, Kulik J, Hamberger M. The brain's orienting response (novelty P3) in patients with unilateral temporal lobe resections. *Neuropsychologia*. 2011; 49(12):3474–3483.10.1016/j.neuropsychologia.2011.08.023 [PubMed: 21906606]
- Goodin DC, Squires KC, Henderson BH, Starr A. Age-related variations in evoked potentials to auditory stimuli in normal human subjects. *Electroencephalography and clinical Neurophysiology*. 1978; 44:447–458. [PubMed: 76553]
- Gurland B, Golden RR, Teresi JA, Challop J. The SHORT CARE: An efficient instrument for the assessment of depression, dementia and disability. *Journal of Gerontology*. 1984; 39:166–169. [PubMed: 6699370]
- Iwaki N, Matsushima H, Kodaira K. Hypercorrection of high confidence errors in lexical representations. *Perceptual and Motor Skills*. 2013; 117:219–235.
- Jacoby, LL.; Hessels, S.; Bopp, KL. Proactive and retroactive effects in memory performance: Dissociating recollection and accessibility bias. In: Roediger, JSNHL., III; Neath, I.; Surprenant, AM., editors. *The nature of remembering: Essays in honor of Robert G Crowder*. Washington, DC: American Psychological Association; 2001. p. 35–54.

- Knight R. Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography and clinical Neurophysiology*. 1984; 59:9–20. [PubMed: 6198170]
- Kulhavy RW, Yekovich FR, Dyer JW. Feedback and response confidence. *Journal of Educational Psychology*. 1976; 68:522–528.
- McIntyre JS, Craik FIM. Age differences in memory for item and source information. *Canadian Journal of Psychology*. 1987; 41:175–192. [PubMed: 3502895]
- Metcalfe J. Correcting errors with true feedback versus overwriting correct answers with errors. *Journal of Experimental Psychology General*. (under revision).
- Metcalfe J, Butterfield B, Habeck C, Stern Y. Neural correlates of people's hypercorrection of their false beliefs. *Journal of Cognitive Neuroscience*. 2012; 24:1571–1583. [PubMed: 22452558]
- Metcalfe J, Finn B. People's hypercorrection of high confidence errors: did they know it all along? *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2011; 37(2):437–48.
- Metcalfe J, Finn B. Hypercorrection of high confidence errors in children. *Learning and Instruction*. 2012; 22:253–261.
- Metcalfe J, Miele DB. 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*. 2014; 3(3):189–197.
- Mitchell DB. How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1989; 15:31–49.
- Paller KA, Kutas M, Mayes AR. An investigation of neural substrates of memory encoding in man. *Psychophysiology*. 1985; 22:607.
- Perlmutter M. What is memory aging the aging of? *Developmental Psychology*. 1978; 14:330–345.
- Ranganath C, Rainer G. Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*. 2003; 4(3):193–202.
- Ranganath C, Yonelinas AP, Cohen MX, Dy CJ, Tom SM, D'Esposito M. Dissociable Correlates of Recollection and Familiarity within the Medial Temporal Lobes. *Neuropsychologia*. 2003; 42:2–13. [PubMed: 14615072]
- Ruch FL. The differentiative effects of age upon human learning. *Journal of General Psychology*. 1934; 11:261–286.
- Semlitsch HV, Anderer P, Schuster P, Presslich O. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*. 1986; 23(6):695–703. [PubMed: 3823345]
- Sharbrough, F.; Chatrian, GE.; Lsner, RP.; Luders, H.; Nuwer, M.; Picton, TW. Guidelines for Standard Electrode Position Nomenclature. Bloomfield: American EEG Society; 1990.
- Sitzman DM, Rhodes MG, Tauber SK, Liceralde VR. The role of prior knowledge in error correction for younger and older adults. *Aging, Neuropsychology and Cognition*. 2015; 22(4):502–516.10.1080/13825585.2014.993302.
- Staudinger UM, Cornelius SW, Baltes PB. The aging of intelligence: Potentials and limits. *Annals of the American Academy of Political and Social Science*. 1989; 503:43–59.
- Tulving E. Episodic Memory From Mind to Brain. *Annual Review of Psychology*. 2002; 53:1–25.
- Umanath S, Marsh EJ. Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*. 2014; 9:408–426. [PubMed: 26173273]

Figure 1. Probability correct on the final surprise test for items that were high and low-confidence errors on the initial semantic task. Error bars indicate ± 95 percent CI's. The data have been averaged across participants in each age group.

Figure 2.

Left column. Grand mean ERPs at FCz averaged across participants within each Age Group (young, top; older, bottom) elicited by the corrective feedback to high- (red lines) and low- (black lines) confidence errors. *Middle column.* Maps of the scalp distribution of P3a averaged voltages associated with the waveforms in the first column. Note that the fronto-central distribution of the P3a is observed in both young and older adults. Calibration bar indicates the magnitude of positive (unshaded) and negative (shaded) amplitudes of the P3a. Dots indicate the electrode locations. *Right column.* Grand mean P3a voltages (based on the waveforms in the left column) averaged across participants in each Age Group to high- and low-confidence errors. Error bars indicate ± 95 percent CI's.

Figure 3. Probability correct on the final surprise test for items that were rated familiar and unfamiliar on the initial semantic task. Error bars indicate ± 95 percent CI's. The data have been averaged across participants in each age group.

Table 1Grand Mean ($\pm$ SD) Demographic Data for Young and Older Adults.

	Young ($\pm$ SD) *	Older ($\pm$ SD)	P value
Age	24.2 (3.1)	73.7 (6.2)	
Education	15.5 (1.6)	16.8 (2.2)	<.002
Modified Mini-Mental	54.6 (2.1)	54.2 (1.9)	ns
IQ (verbal)	118.7 (14.1)	117.2 (7.6)	ns
IQ (performance)	107.9 (14.9)	114.7 (10.8)	<.02
Digit Forward	7.2 (1.3)	6.8 (1.3)	ns
Digit Backward	5.1 (1.3)	5.2 (1.2)	ns
Beck Depression	5.7 (5.2)	3.9 (3.4)	ns
SHORT CARE (Gurland, Golden, Teresi, & Challop, 1984)			
Dementia (cutoff = 6)	NA	0.3 (0.5)	
Depression (cutoff = 7)	NA	2.2 (1.4)	

ns = not significant; NA = not applicable

* Two young participants are missing demographic data.

Table 2

Sequence for a Single Trial

1. "Please Wait" Display (experimenter starts trial)
2. Pre question Interval (800–1200 ms variable delay)
3. Question (participant has unlimited time to respond) When participant responds aloud, experimenter records response by typing and marks its correctness; the confidence screen is then presented
4. Confidence screen (participant has unlimited time to mark confidence on keyboard from 1–7)
5. Delay (500 ms)
6. Tone (300 ms) to alert participant that feedback will be displayed
7. Delay (1500 ms)
8. Feedback (500 ms) All ERP recordings begin at the feedback signal; the correct answer is displayed on the computer screen in green if participant's response is correct, red if incorrect; familiarity screen is then displayed
9. Familiarity (participant has unlimited time to respond). If the participant's response is incorrect, participant enters their familiarity with the correct response on the keyboard (1 = familiar; 2 = unfamiliar). If their response is correct, skip familiarity and start new trial.
During the re-test phase, confidence and familiarity screens are not displayed.