Plasticity of Inhibition in Older Adults: Retest Practice and Transfer Effects

Andrea J. Wilkinson and Lixia Yang
Ryerson University, Toronto, Ontario, Canada

This study assessed plasticity of inhibition in older adults through examining retest practice effects in a six-session training paradigm using the Stroop task and the training-induced transfer effects to a range of cognitive measures. Fifty-six older adults (aged 60 – 84 years, mean = 71.05, standard deviation = 6.17) participated in this study. They were evenly assigned to one of four groups: summary feedback, individualized and adaptive feedback, no-feedback, or a no-contact control group. The results suggest that older adults are able to improve inhibition across retest sessions but unable to transfer the retest practice effects to other tasks. In addition, the improvement is not item specific and feedback does not appear to moderate the magnitude of the training benefits; however, feedback does appear to reduce interference variance across retest practice sessions.

Keywords: aging, inhibition, retest practice, transfer, feedback

Cognitive plasticity refers to the ability to improve cognitive performance through training (Baltes & Lindenberger, 1988). A substantial body of literature suggests that age-related cognitive declines in older adults can be slowed down or even reversed through cognitive training (for reviews, see Baltes & Lindenberger, 1988; Greenwood, 2007; Hertzog, Kramer, Wilson, & Lindenberg, 2008; Kramer & Willis, 2002; Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010; Lustig, Shah, Seidler, & Reuter-Lorenz, 2009; Stine-Morrow & Basak, 2011; Thompson & Foth, 2005). Training-induced plasticity in older adults has been demonstrated in a variety of functional domains such as reasoning, processing speed, memory, and executive functioning (e.g., Ball et al., 2002; Baltes, Sowarka, & Kliegl, 1989; Kramer, Hahn, & Gopher, 1999; Yang, Krampe, & Baltes, 2006).

Recent studies suggest that older adults maintain plasticity in executive functions such as dual-task performance (Bherer et al., 2005), working memory updating (Dahlin, Nyberg, Bäckman, & Neely, 2008), and task switching (Kramer et al., 1999). According to the executive function model proposed by Miyake, Friedman, Emerson, Witzki, and Howertor (2000), inhibition is one of the key executive functions (in addition to updating and task-switching). Inhibition serves to control the contents of working memory (Hasher, Zacks, & May, 1999). A large body of empirical studies suggests that older adults have persistent deficits in inhibition (Bojko, Kramer, & Peterson, 2004; Bulter & Zacks, 2006; Gazzaley, Cooney, Rissman, & D’Esposito, 2005; Hamm & Hasher, 1992; Nieuwenhuis, Ridderinkhof, de Jong, Kok, & van der Molen, 2000; Rowe, Valderrama, Hasher, & Lenartowicz, 2006; Yang & Hasher, 2007). According to the inhibitory deficit hypothesis of aging (Hasher et al., 1999), many age-related cognitive declines are due to older adults’ deficits in inhibition. In support of this theory, it has been shown that inefficient inhibition in older adults led to deficits in other cognitive domains such as speed of processing (Lustig, Hasher, & Tonev, 2006) and memory (Gazzaley et al., 2005).

Despite the importance of inhibition in cognitive functions of the aging brain, little research has been done to examine the plasticity of inhibition in older adults. To our knowledge, two extant studies revealed encouraging findings that older adults were able to improve their inhibition performance across trials within one or two sessions of practice (Davidson, Zacks, & Williams, 2003; Dulaney & Rogers, 1994). However, there are some methodological limitations in these studies. First, they used the same color word stimuli throughout all of the trials within a given session (e.g., blue, green, red, yellow, and white), and thus the improvement could be readily explained by item-specific effects (i.e., knowledge or skill acquired in naming the specific colors; Yang, Reed, Russo, & Wilkinson, 2009). Second, the extensive practice within one or two sessions is particularly vulnerable to cumulative fatigue and may thus limit performance improvement. For example, Dulaney and Rogers (1994) found that older adults did not develop an automatic reading suppression response within the one-session practice on the Stroop task and they argued the performance improvement in older adults was mainly driven by task-related factors such as efficiency of naming specific colors and improved visual scanning. Third, mass practice does not allow much opportunity for repeated testing, which has been shown to be particularly beneficial for long-term memory (Karpicke & Roedi-
In a typical Stroop task, there are three different types of trials: congruent (e.g., the word “BLUE” printed in blue ink, respond blue); incongruent (e.g., the word “BLUE” printed in green ink, respond green); and neutral trials (e.g., “XXXX” printed in blue ink, respond blue; MacLeod, 1991, following Stroop, 1935). Empirical data have consistently demonstrated that responses were slower and more prone to errors on incongruent relative to neutral trials. This effect is labeled as the Stroop interference effect, and it is thought to arise from the conflict between two incoming sources of information or responses (i.e., word reading and color naming) and the inability to keep the irrelevant information (i.e., color word) outside of working memory (Cohen, Dunbar, & McClelland, 1990; Faust & Balota, 2007; Jonides, Badre, Curtis, Thompson–Schill, & Smith, 2002; MacLeod, 1991; MacLeod, 1998; Milham et al., 2002; Stroop, 1935). As such, it is commonly assumed that the Stroop interference effect is driven by deficits in inhibition of irrelevant “color word” information (Faust & Balota, 2007; Milham et al., 2002; West & Alain, 2000). Given the inhibitory deficits in older adults, it is not a surprise that they consistently demonstrate larger Stroop interference effects when compared to younger adults (e.g., Davidson et al., 2003; Dulaney & Rogers, 1994; Muter, Naylor, & Patterson, 2005). Therefore, practice on the Stroop task provides a valid measure for the plasticity of inhibition.

Transfer effects assess the generalizability of training-induced improvement in the trained tasks to other untrained tasks. They are most likely to be evidenced in tasks that overlap with the trained task in the basic underlying cognitive processes (Thorndike & Woodworth, 1901). Transfer effects have been distinguished into either near or far transfer effects (i.e., transfer to tasks that are overlapping or not overlapping in underlying processes tapped by the trained tasks, respectively; Dahlin et al., 2008). Research findings are mixed in terms of transfer effects in executive training with older adults. Though some studies suggest that older adults show near transfer effects in training with dual-task performance (Bherer et al., 2005) and task-switching (Karbach & Kray, 2009), other studies fail to reveal near transfer effects in updating training (e.g., Dahlin et al., 2008). Furthermore, older adults show little far transfer effects, except for in one study, which used task-switching training (Karbach & Kray, 2009). The current study aims to determine whether older adults are able to transfer practice effects on the Stroop task to other untrained cognitive tasks. To evaluate potential transfer effects, we administered a large battery of cognitive tasks, including a near transfer task measuring inhibition, and far transfer tasks assessing other cognitive functions (i.e., speed of processing, reasoning, attention, and task-switching) at pretest and posttest sessions.

In the literature on executive function training, two different types of feedback have been implemented. Specifically, Kramer et al. (1999) presented summary feedback in a task-switching training program. In this study, participants completed blocks of trials and at the end of each block they received a summary feedback display, which presented the average reaction time (RT) and accuracy within that block. With this feedback, younger and older adults showed equal retest practice effects. In addition, Bherer et al. (2005) presented trial by trial, individualized and adaptive feedback in a dual-task training program. In this study, a bar graph was continuously presented throughout the training sessions, and the bar appeared in different colors depending on whether the RT performance on the current trial improved or declined (i.e., faster or slower) compared to the average RT of previous trials. Surprisingly, this training paradigm resulted in greater retest practice effects in older compared to younger adults. Despite the impressive practice effects with these two types of feedback in older adults, no studies have directly examined which feedback is more effective. Moreover, without a no-feedback control, it is unclear whether feedback adds any extra benefits. Some recent studies (Yang et al., 2006; Yang et al., 2009) demonstrated substantial retest practice effects in reasoning, speed, and attention even without any feedback. Given the lack of empirical comparison, the effect of feedback on practice effects in older adults is still an open question. To assess the moderating effect of feedback on the retest practice effect in inhibition, we manipulated and compared summary feedback and individualized and adaptive feedback against a no-feedback control group across the six retest practice sessions.

In summary, the current study addresses the following three questions: 1) Do older adults show plasticity in inhibition through retest practice with a multiple-session single-item Stroop training paradigm? 2) Do older adults show near and/or far transfer effects in Stroop inhibitory training? 3) Does feedback impact plasticity of inhibition (i.e., retest practice and transfer effects) in older adults?

Method

Participants

Fifty-six older adults (20 males and 36 females, age range = 60 – 84 years; mean $M = 71.05$, standard deviation $SD = 6.17$)
were recruited to participate in this study. They were randomly and evenly assigned to the three training groups and one no-contact control group. The three training groups were: summary feedback, individualized and adaptive feedback and no-feedback groups. No participants presented with color blindness, as measured with the Dvorine Pseudo-Isochromatic Plates (Dvorine Pseudo-Isochromatic Plates, 1953), and all had shown to be normal near vision, as measured with the Rosenbaum near acuity pocket screener (Rosenbaum Pocket Vision Screener, 1986). No participants showed dementia related cognitive impairment as screened with the Short-Blessed Test (SBT; Katzman et al., 1983), and none of them reported severe anxiety, as reflected in the scores on the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988). They were compensated $10 for each hour of their participation.

The three feedback groups did not differ in any of the demographic characteristics or baseline cognitive performance scores (ps > .07; see Table 1). Similarly, the training groups and no-contact control group did not differ in most demographic characteristics or baseline cognitive performance scores (ps > .08; see Table 1), except that the no-contact control group had significantly more years of education, and scored higher on Digit Symbol and Letter Series (ps ≤ .02). However, the results on the transfer effects remained the same even after controlling for these differences in covariate analyses.

Materials

A 17-inch monitor PC computer was used for all the computerized tasks at the pretest, posttest, and all of the training sessions. Participants were comfortably seated in a well-lit testing room at a viewing distance of approximately 60 cm for all of the sessions.

Training materials and stimuli. Each of the six training sessions involved the completion of a computerized Stroop task. Stroop stimuli were created from seven colors (i.e., blue, pink, green, brown, yellow, orange, and purple). At each session four colors were used, including two constant colors that were repeated across all sessions, and two varying colors selected from the remaining five colors. The two colors serving as the constant colors were either blue and orange (counterbalance Condition 1) or pink and yellow (counterbalance Condition 2), counterbalanced across participants. At each session, the four color words were matched on word length across the two counterbalance conditions (e.g., at Session 1: both counterbalance conditions included two 4-letter color words, one 5-letter color word, and one 6-letter color word). The pairing of the two varying colors was different at each session within each counterbalance condition. In counterbalance Condition 1, the pair of varying colors was green and pink for Session 1, brown and yellow for Session 2, green and purple for Session 3, yellow and pink for Session 4, purple and brown for Session 5, as well as yellow and green for Session 6. In counterbalance Condition 2, the pair of varying colors was blue and green for Session 1, orange and brown for Session 2, green and purple for Session 3, blue and orange for Session 4, purple and brown for Session 5, as well as orange and green for Session 6.

The stimuli were constructed in such a way that the three trial types (i.e., congruent, incongruent, and neutral) were pseudorandomly mixed within each block, such that the same trial type never appeared more than three times in a row. In addition, there was no possibility for negative priming (i.e., the distracting color word on the current trial was never the ink color on the subsequent trial; Little & Hartley, 2000). Similarly, the ink color on the current trial was never the distracting color word on the subsequent trial.

Pretest and posttest measures. To assess near and far transfer effects and to evaluate baseline cognitive performance, a variety of cognitive measures were administered at pretest and posttest sessions. The near transfer inhibition task was a Go–No Go task.

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Summary feedback (n = 14)</th>
<th>Individualized and adaptive feedback (n = 14)</th>
<th>No feedback (n = 14)</th>
<th>No-contact control (n = 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>70.43 (6.24)</td>
<td>72.71 (6.03)</td>
<td>69.79 (7.05)</td>
<td>71.29 (5.57)</td>
</tr>
<tr>
<td>Education (yrs)</td>
<td>16.39 (3.81)</td>
<td>15.79 (4.00)</td>
<td>15.88 (3.59)</td>
<td>18.14 (2.51)</td>
</tr>
<tr>
<td>Health</td>
<td>8.84 (1.01)</td>
<td>8.43 (1.02)</td>
<td>8.57 (1.02)</td>
<td>8.43 (0.54)</td>
</tr>
<tr>
<td>Visual Acuity (20/–)</td>
<td>30.71 (5.84)</td>
<td>36.43 (14.20)</td>
<td>30.36 (8.65)</td>
<td>27.14 (7.26)</td>
</tr>
<tr>
<td>SBT</td>
<td>0.71 (1.68)</td>
<td>1.43 (1.83)</td>
<td>0.64 (1.08)</td>
<td>0.57 (0.94)</td>
</tr>
<tr>
<td>BAI</td>
<td>7.79 (6.04)</td>
<td>7.64 (6.32)</td>
<td>5.36 (6.46)</td>
<td>4.64 (3.65)</td>
</tr>
<tr>
<td>Shipley</td>
<td>36.29 (4.68)</td>
<td>36.00 (3.21)</td>
<td>36.50 (3.78)</td>
<td>36.71 (3.23)</td>
</tr>
<tr>
<td>Go–No Go</td>
<td>8.21 (4.42)</td>
<td>7.14 (5.46)</td>
<td>4.57 (2.06)</td>
<td>6.86 (3.30)</td>
</tr>
<tr>
<td>DS</td>
<td>61.64 (16.16)</td>
<td>58.57 (11.93)</td>
<td>59.86 (10.51)</td>
<td>70.64 (17.19)</td>
</tr>
<tr>
<td>LS</td>
<td>9.71 (4.94)</td>
<td>10.00 (3.53)</td>
<td>8.50 (4.69)</td>
<td>12.71 (4.27)</td>
</tr>
<tr>
<td>ANT – Alerting</td>
<td>18.98 (29.43)</td>
<td>34.45 (32.72)</td>
<td>33.87 (53.97)</td>
<td>31.94 (39.38)</td>
</tr>
<tr>
<td>ANT – Orienting</td>
<td>76.55 (47.56)</td>
<td>81.13 (35.15)</td>
<td>69.87 (40.53)</td>
<td>73.67 (27.13)</td>
</tr>
<tr>
<td>ANT – Executive Control</td>
<td>116.46 (36.93)</td>
<td>105.39 (63.56)</td>
<td>110.07 (64.53)</td>
<td>130.16 (50.11)</td>
</tr>
<tr>
<td>Task-switching</td>
<td>−12.67 (277.48)</td>
<td>34.05 (190.45)</td>
<td>43.80 (176.98)</td>
<td>56.47 (153.85)</td>
</tr>
</tbody>
</table>

Note. Mean score with standard deviations presented in parentheses. Health = health rating on a self-report scale (average score, out of 10); Visual acuity = near visual acuity score from the Rosenbaum near visual acuity pocket screen (SBT = Short-Blessed Test (average score); BAI = Beck Anxiety Inventory (average score); Shipley = Shipley Vocabulary Test (average number correct, out of 40); Go–No Go (average number of false alarms); DS = digit symbol (average number of correct solutions); LS = letter series (average number of correct solutions, out of 20); ANT = Attention Network Test (difference scores in ms); Task-Switching = difference score in ms.
(Donders, 1868/1969). The far transfer tasks included the Digit Symbol (DS; Wechsler, 1981) as a measure of perceptual speed, Letter Series (LS; Blieschner, Willis, & Baltes, 1981) as a measure of inductive reasoning, the Attention Network Test (ANT; Fan, McCandliss, Sommer, Raz, & Posner, 2002) as a measure of attention, and a Task-Switching task (Kumada et al., 2005) as a measure of task switching.

Procedure

Participants in the training groups completed one 1.5-hr pretest session, six 30-min training sessions (spread over 2 weeks), and one 1-hr posttest session. Each session was scheduled individually at the most convenient time slot for each participant that fit in the general training schedule framework: three times per week for two consecutive weeks. Upon completion of the whole procedure, participants were paid and debriefed.

Training sessions. Each training session involved four blocks of retest practice on the Stroop task in the following fixed sequence: Key-Color Acquisition, Practice, Training, and Standard. The Practice, Training, and Standard blocks (i.e., Blocks 2–4, respectively) all included color words and X strings printed in colored ink. But, in the Key-Color Acquisition block (i.e., Block 1), only X strings were used. In the Stroop task, participants were instructed to identify and respond to the ink color by pressing the correspondingly assigned color-coded response keys on the Stimulus–Response (S–R) box. The key-color mapping assignment was counterbalanced across sessions and participants.

Block 1: Key-color acquisition. This block consisted of 40 neutral trials (i.e., X strings printed in colored ink) that served to help participants associate the S-R box keys with the assigned Stroop colors. The number of Xs in each X string was the same as the number of letters in the corresponding ink color name (e.g., XXXX in pink colored ink, XXXXX in green ink, and XXXXXX in orange ink etc.) Each trial began with a fixation cross (+) presented at the center of the screen for 1000 ms, which was then replaced by a neutral Stroop stimulus (e.g., XXXX in blue ink). Each Stroop stimulus was presented for up to 5000 ms or terminated by a key-press response. Following each response, a feedback screen (presented for 1500 ms) appeared, indicating whether the response was “Correct!” or “Incorrect,” before proceeding to the next trial.

Block 2: Practice. This block consisted of 24 trials (eight for each trial type: congruent, incongruent, and neutral). The purpose of the practice block was to familiarize participants with the trial procedure for the upcoming training block (i.e., Block 3). The trial procedure paralleled that of Block 1, except that the individualized and adaptive feedback group received the trial-by-trial feedback manipulation during this block. The feedback display was presented for 2000 ms and it provided feedback on accuracy and RT on the current trial, as well as a bar graph, which depicted relative performance. The color and height of the bar graph was determined by the RT of the current trial relative to the average RT of all previous trials in the block (i.e., RT of current trial—average RT of all previous trials). Specifically, the bar graph would turn green, if the current RT was faster, and turn red, if the current RT was slower, than the average RT of all previous trials. Also, the height of the bar graph was determined by this RT difference. If an incorrect response was made, a feedback screen displayed “Incorrect” for 2000 ms. In contrast, the summary feedback and no-feedback groups did not receive any trial by trial feedback during this block to mimic the trial procedure for the upcoming training block (i.e., Block 3).

Block 3: Training. The training block consisted of 216 trials (72 for each trial type: congruent, incongruent, and neutral). The trial procedure was the same as in the practice block, except for the addition of a summary feedback display that was presented at the end of the block for the summary feedback and individualized and adaptive feedback groups. The summary feedback display consisted of a screen showing mean RT, and percentage of correct responses for the block, and a message which reminded participants to always try to improve speed and accuracy at each training session. The no-feedback group served as a control, and thus received no feedback during this block.

Block 4: Standard. The standard block consisted of 144 trials (48 for each trial type: congruent, incongruent, and neutral), and was designed to provide a fair comparison between the three training groups. For this purpose, no feedback was provided to any of the three training groups. Other than that, the trial procedure remained the same as in the training and the practice block.

In total, there were 424 trials (128 congruent, 128 incongruent and 168 neutral) presented at each training session. In the critical experimental block (i.e., Standard, Block 4) used for data analysis, 144 trials were presented (48 of each trial type) in which each of the four chosen colors served as the ink color 12 times in each of the three trial types (i.e., congruent, incongruent, and neutral) at a given session. For the incongruent trials, each color was presented an equal number of times as the color word and the ink color of mismatched color words. For example, the word “BLUE” appeared 12 times as the color word, and blue appeared as the ink color four times for each of “PINK,” “GREEN,” and “YELLOW” color words (i.e., 12 times in total).

Results

All statistical analyses were conducted using SPSS 17.0. In the following section, we report the results on retest practice effect (i.e., performance improvement across the six retest training sessions) and then on transfer effects.

Retest Practice Effects

All dependent variables were calculated on the data collected during the standard block (i.e., Block 4) to control for variation in the block procedure across the three training groups. The analyses were conducted on the RT and accuracy (i.e., proportion of correct responses). Only RTs for correct responses were included in the RT analysis. RTs were trimmed such that any RTs beyond 2.5 SDs from the mean for each trial type at each session were excluded from the final data analysis. As a result, 2.6% of the RTs were deleted. Due to technical problems, one subject did not complete the standard block at the first session. The missing RT and accuracy data were estimated using the linear regression model based on the performance of this participant at the other five completed sessions (i.e., Session 2 through 6).

Overall RT and accuracy. Preliminary mixed Model 3 (trial type) × 3 (feedback) × 6 (session) analyses of variance (ANOVAs) were conducted on RT and accuracy. Session and trial
type were within-subjects variables, whereas feedback was a between-subjects variable. To best capture the nature and trajectory of the training benefits, the linear (suggesting incremental improvement) and quadratic contrasts (reflecting saturation effects) were specified for the session effect. See Table 2 for the overall RT and accuracy scores at each session separately for the three feedback groups.

The overall RT analysis revealed a significant session effect in the linear trend. $F(1, 39) = 134.00$, $p < .001$, $\eta^2 = .78$, and approaching significant effect in the quadratic trend, $F(1, 39) = 3.85$, $p = .06$, $\eta^2 = .09$. Furthermore, there was a main effect of trial type, $F(2, 78) = 92.22$, $p < .001$, $\eta^2 = .70$. Post hoc $t$ tests revealed that overall RT of incongruent trials ($M = 967.00$, $SD = 248.87$) was significantly greater than both congruent ($M = 818.03$, $SD = 180.23$) and neutral trials ($M = 823.14$, $SD = 173.08$), $p < .001$, the latter two did not differ significantly ($p = .29$). All the significant effects in the post hoc comparisons remained significant after applying the Bonferroni corrections for three multiple comparisons to adjust $p$ values ($ps < .003$). The main effect of feedback was not significant ($p = .35$). The significant linear session effect was qualified by a significant session by trial-type interaction, $F(2, 82) = 12.32$, $p < .001$, $\eta^2 = .23$. To further explore this interaction, individually estimated linear slopes were calculated separately for each trial type and submitted to paired $t$ tests analyses. The results indicated that the slope for incongruent trials ($M = -32.17$, $SD = 21.93$) was significantly steeper than both congruent ($M = -19.21$, $SD = 13.04$), and neutral trials ($M = -19.64$, $SD = 15.58$), $p < .001$, the latter two did not differ significantly ($p = .81$). No other effects were significant, $ps > .14$.

The overall accuracy analyses indicated that there were no significant linear or quadratic effects of session ($ps > .28$). However, the main effect of trial type was significant, $F(2, 78) = 6.05$, $p = .004$, $\eta^2 = .13$. Post hoc paired $t$ tests revealed that overall accuracy of incongruent trials ($M = .96$, $SD = .05$) was significantly lower than both congruent ($M = .98$, $SD = .02$) and neutral trials ($M = .98$, $SD = .02$), $ps < .01$, the latter two did not differ significantly ($p = .18$). All the significant effects in the post hoc comparisons remained significant after applying the Bonferroni corrections for three multiple comparisons to adjust $p$ values ($ps < .051$). The session by trial type interaction was approaching, but did not reach significance ($Fs < 2.98$, $ps > .06$). The main effect of feedback failed to reach significance ($p = .07$). No other effects were significant, $ps > .47$.

The results in the overall RT, but not in accuracy, suggested that Stroop performance was greatly improved, with a differentially greater improvement in incongruent trials relative to both neutral and congruent trials, across the six retest training sessions. This implied that the interference, primarily presented in the incongruent trials, may be greatly reduced across sessions. To specifically test this, we conducted analyses on the Stroop interference scores calculated on RT.

### Stroop interference scores (RT)

To specifically assess the efficiency of inhibition, we calculated the interference scores as the difference in RT between incongruent and neutral trials (i.e., RT interference score = incongruent RT – neutral RT). The preliminary analyses showed no significant differences between congruent and neutral trials in both RT and accuracy, suggesting an absence of Stroop facilitation effects. As such, no further analyses on facilitation were conducted.

The Stroop interference scores were submitted to a mixed-model ANOVA, with session as a within-subjects variable and feedback as a between-subjects variable. The analysis revealed a significant linear session effect, $F(1, 39) = 11.86$, $p = .001$, $\eta^2 = .23$ (see Figure 1), suggesting that the RT interference scores were gradually reduced across the retest training sessions. No other effects were significant, $ps > .15$.

Although the feedback did not moderate the mean performance improvement across sessions, it may have affected training-related change in between-individuals performance variance. For this purpose, we calculated the SDs of RT interference scores for each feedback group (i.e., summary feedback, individualized and adaptive feedback, and no feedback) at each session. For the summary feedback group, the SDs at Session 1 to Session 6 were 267.09 ms, 111.38 ms, 109.44 ms, 77.02 ms, 66.63 ms, and 118.60 ms, respectively, with a linear slope of $-25.98$. For the individualized and adaptive feedback group, the SDs at Session 1 to Session 6 were 93.76 ms, 72.45 ms, 93.49 ms, 65.05 ms, 37.18 ms, and 52.66 ms, respectively, with a linear slope of $-9.71$. Finally, for the no-feedback group, the SDs at Session 1 to Session 6 were 138.54 ms, 174.65 ms, 104.39 ms, 141.97 ms, 123.38 ms, and 190.00 ms, respectively, with a linear slope of 4.03. The linear regression model analyses (with session as the predictor) revealed that the linear model was not significant for the summary feedback ($p = .15$) and the no-feedback group ($p = .65$); however, just reached significant for the individualized and adaptive feedback group ($p = .05$). We should note that both feedback groups showed a

![Table 2](Image)

The Overall RT (in Ms) and Accuracy (Proportion) for the Training Groups at Session 1–6

<table>
<thead>
<tr>
<th>Session</th>
<th>Summary feedback (n = 14)</th>
<th>Individualized and adaptive feedback (n = 14)</th>
<th>No feedback (n = 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>Accuracy</td>
<td>RT</td>
</tr>
<tr>
<td>1</td>
<td>917.12 (284.40)</td>
<td>.96 (.07)</td>
<td>919.45 (117.45)</td>
</tr>
<tr>
<td>2</td>
<td>899.92 (289.43)</td>
<td>.96 (.05)</td>
<td>820.96 (106.37)</td>
</tr>
<tr>
<td>3</td>
<td>891.80 (282.05)</td>
<td>.97 (.03)</td>
<td>820.15 (109.17)</td>
</tr>
<tr>
<td>4</td>
<td>844.54 (306.91)</td>
<td>.97 (.03)</td>
<td>789.12 (96.94)</td>
</tr>
<tr>
<td>5</td>
<td>813.73 (239.29)</td>
<td>.97 (.03)</td>
<td>782.08 (105.25)</td>
</tr>
<tr>
<td>6</td>
<td>837.86 (283.42)</td>
<td>.97 (.03)</td>
<td>762.33 (114.88)</td>
</tr>
</tbody>
</table>

**Note.** Mean scores with standard deviations presented in parentheses.
negative slope (though the linear model was only significant for the individualized and adaptive feedback group), whereas only the no-feedback group showed a positive, though nonsignificant, slope in interference variance across sessions. In sum, feedback did not moderate the retest-induced performance improvement (i.e., mean interference reduction); however, it appeared to have reduced the interference variance across retest sessions.

**Item-specific effects.** To assess the contribution of item-specific effects, we calculated RT interference scores separately for constant and varying color stimuli. The resulting interference scores were then submitted to a mixed model ANOVA, with session and color type (constant vs. varying) as within-subject variables and feedback (summary feedback vs. individualized and adaptive feedback vs. no feedback) as a between-subjects variable. Replicating the overall analysis, there was a significant linear session effect, $F(1, 39) = 19.95, p < .001, \eta^2 = .34$, indexing a gradual reduction in RT interference scores across retest sessions. There was also a significant effect of color type, $F(1, 39) = 25.26, p < .001, \eta^2 = .39$, with a greater RT interference score for the constant colors ($M = 163.71, SD = 100.87$) than for the varying colors ($M = 128.98, SD = 93.47$). However, the session (in linear or quadratic trends) by color type interactions were not significant ($ps > .18$), suggesting equivalent performance gains for both constant and varying colors. This implies that the retest improvement in Stroop performance is not driven by the item-specific effects of improving stimulus-specific processing efficiency, but instead is likely the result of general improvement in processing efficiency to inhibit/suppress irrelevant word information. No other effects were significant, $ps > .09$. Taken together, the results on retest practice effects indicate a significant reduction in Stroop interference scores (calculated with RT) across sessions, which is likely driven by generally improved inhibition instead of an item-specific effect.

**Transfer Effects**

In order to investigate transfer effects, we initially ran separate 2 (session: pretest vs. posttest) $\times$ 3 (feedback: summary feedback vs. individualized and adaptive feedback vs. no feedback) mixed-model ANOVAs on the outcome measures of inhibition, speed of processing, reasoning, attention, and task switching administered at pretest and posttest sessions. These preliminary analyses did not reveal any significant session by feedback interactions (all $ps > .29$), indicating that all of the three feedback groups showed the same pretest–posttest improvement. As such, for all subsequent analyses, we collapsed across feedback groups to create an overall “training” group and conducted separate 2 (session: pretest vs. posttest) $\times$ 2 (group: training vs. no-contact control) mixed-model ANOVAs for each outcome measure administered at pretest and posttest sessions. Table 3 displays the pretest and posttest performance of the two groups and the $p$ values for the interactions.

**Near transfer effect.** The number of false alarms in the Go–No Go task (Donders, 1868/1969) was used to assess the near transfer effect. Despite a reduction in false alarms from pretest to posttest sessions ($p = .01$), the session by group interaction was not significant ($p = .93$), indicating the absence of near transfer effect.

**Far transfer effects.** The average number of correct items was used as the dependent variable for the DS (as a measure of perceptual speed) and the LS (as a measure of inductive reasoning). In the ANT (as a measure of attention), RT difference scores were used to measure three different subtypes of attention (i.e., alerting, orienting, and executive control; see Fan et al., 2002 for details on the calculation methods), with larger difference scores reflecting poorer attention performance. Finally, in the task-switching task, RTs on nonswitch trials (i.e., rule A–rule B) were subtracted from switch trials (i.e., rule A–rule B) and this difference score was used as a measure of task-switching ability. The ANOVAs on the far transfer outcome measures showed a significant pretest–posttest performance improvement in all of the tasks except for task-switching ($ps < .03$). However, the critical interactions were not significant ($ps > .21$), suggesting the absence of far transfer effects from Stroop inhibition training.

**Discussion**

The current study aimed to examine plasticity, evaluated by the retest practice and transfer effects, of inhibition using a 6-session single-item Stroop training paradigm. The results suggest that (1) older adults maintain plasticity of inhibition, as indexed by a significant reduction in Stroop interference scores across retest sessions; (2) there are no near or far transfer effects to other selected cognitive measures; (3) feedback does not appear to moderate the magnitude of inhibition plasticity, despite a tendency to reduce interindividual interference variance in RTs across retest sessions.

**Retest Practice Effect**

The current study demonstrated that older adults could improve the efficiency of inhibition with retest training. Specifically, Stroop RT interference scores were reduced across six retest sessions and the benefit was the same for constant and varying colors. Going beyond the existing literature on the plasticity of inhibition in older adults measured with single-session practice on a repeated-color Stroop paradigm (e.g., Dulaney & Rogers, 1994), the current study adds strong and novel evidence suggesting that inhibition can improve robustly with a multiple-session retest training paradigm for older adults, and the improvement induced in
this training paradigm is likely driven by developing general inhibition efficiency instead of item-specific processing. Such a finding is consistent with previous work showing reliable retest learning in the absence of item-specific effects for older adults in reasoning, speed, and attention (Yang et al., 2009). Furthermore, the demonstrated plasticity of inhibition in older adults extends literature on the plasticity of executive functions in older adults, as shown previously in dual task (e.g., Dahlin et al., 2005; Edwards et al., 2005), updating (e.g., Dahlin et al., 2008), and task switching (e.g., Karbach & Kray, 2009; Kramer et al., 1999).

In literature, Stroop interference, but not facilitation, has typically been used as a measure of inhibition (e.g., Dulaney & Rogers, 1994; Little & Hartley, 2000; Verhaeghen & De Meersman, 1998; West & Alain, 2000). The current study extends the earlier findings of reduced Stroop interference effects across practice blocks within a single or two sessions (Davidson et al., 2003; Dulaney & Rogers, 1994) in older adults in a systematically manipulated multiple-session training paradigm, in the absence of item-specific effects. As in Dulaney et al. (2003), the ratio of errors committed was very low in the present study. Therefore, the lack of improvement in accuracy interference scores is likely due to ceiling effects. Specifically, when collapsed across feedback groups, the biggest drop in accuracy interference scores occurred from Session 1 (4.2%) to Session 2 (1.6%), after which accuracy interference scores hovered around 1.0% for the remaining sessions, leaving little room for further improvement.

Near and Far Transfer Effects

The transfer effect analyses showed neither near nor far transfer effects of Stroop inhibition training in older adults. The lack of far transfer effects was not surprising given substantial literature showing no far transfer effects from cognitive training with older adults (e.g., Dahlin et al., 2008). To our knowledge, only one study, on task-switching training, showed far transfer in older adults (Karbach & Kray, 2009). However, the lack of near transfer effect in older adults was somewhat unexpected, given that near transfer effects have previously been demonstrated in older adults following executive functioning training (e.g., Bherer et al., 2005; Karbach & Kray, 2009). We offer some speculations for the discrepancy. First, it may be due to the type of task used to assess the near transfer effect in the current study. Hasher and colleagues (1999) proposed three subfunctions of inhibition (i.e., access, deletion, and restraint). It is possible that the type of inhibition being utilized in the near transfer task (i.e., Go–No Go) was fundamentally different from the type of inhibition required in the training task (i.e., Stroop). It is possible that the Stroop task is exercising the access function of inhibition, which requires keeping irrelevant information outside the focus of attention, while the Go–No Go task is targeting the restraint subfunction, which requires withholding an automatic response that is inappropriate for the task at hand (Hasher et al., 1999). Second, executive function training studies that have demonstrated near transfer effects in older adults (e.g., Bherer et al., 2005; Karbach & Kray, 2009) utilized the exact same tasks for training and transfer (varying only the specific items). In contrast, for those studies suggesting no near transfer effects in older adults (e.g., Dahlin et al., 2008), tasks used at training were different from those used to assess transfer effects, despite that they may assess similar underlying cognitive processes. Therefore, it is also possible that while the Stroop and Go–No Go task tap an overlapping cognitive process (i.e., inhibition), older adults have limited capacity to evoke near transfer effects to a different task. Nevertheless, we need to be cautious to avoid overgeneralizing our results. The lack of transfer effects may be due to the specific tasks chosen to measure the transfer effects.

The Effects of Feedback

The current study found that the three feedback groups demonstrated comparable reductions in RT interference scores with retest practice, suggesting similar improvements in inhibitory processing. To rule out the possibility that the lack of feedback effect on the retest practice benefit was due to low statistical power, considering the small sample size for each feedback group, we ran a post hoc power analysis using GPower software (Erdfelder, Faul, & Buchner, 1996) for the within-between interaction (i.e., session by feedback interaction in this study) in the repeated-measures ANOVA. It revealed a high statistical power of 0.997, given the current sample size and design, to detect a medium effect (Cohen effect size index f = .25).

The lack of feedback effects in the retest-practice-induced decrease in interference is in line with earlier findings that older

### Table 3

Performance on the Transfer Cognitive Measures at Pretest and Posttest for the Training and No-Contact Control Groups

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Training (n = 42)</th>
<th>No-Contact Control (n = 14)</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
</tr>
<tr>
<td>Go–No Go</td>
<td>6.64 (4.41)</td>
<td>5.38 (3.45)</td>
<td>6.86 (3.30)</td>
</tr>
<tr>
<td>DS</td>
<td>60.02 (12.83)</td>
<td>65.17 (16.44)</td>
<td>70.64 (17.19)</td>
</tr>
<tr>
<td>LS</td>
<td>9.40 (4.37)</td>
<td>10.57 (4.37)</td>
<td>12.71 (3.30)</td>
</tr>
<tr>
<td>ANT – Alerting</td>
<td>29.10 (39.87)</td>
<td>27.17 (27.91)</td>
<td>31.94 (39.38)</td>
</tr>
<tr>
<td>ANT – Orienting</td>
<td>75.85 (40.64)</td>
<td>73.34 (52.54)</td>
<td>73.67 (27.13)</td>
</tr>
<tr>
<td>ANT – Executive Control</td>
<td>110.64 (55.27)</td>
<td>80.81 (53.18)</td>
<td>130.16 (50.11)</td>
</tr>
<tr>
<td>Task-Switching</td>
<td>21.72 (215.56)</td>
<td>39.60 (200.86)</td>
<td>56.47 (153.85)</td>
</tr>
</tbody>
</table>

Note. Mean scores with standard deviations presented in parentheses. The p-values presented are for the Session (pre vs. post) X Group (training vs. no-contact control) interactions and represent the significance of transfer effects. Go–No Go (average number of false alarms); DS = digit symbol (average number of correct solutions); LS = letter series (average number of correct solutions, out of 20); ANT = Attention Network Test (difference scores in ms); Task-Switching = difference scores in ms.
adults can improve their performance without any externally provided performance monitoring or feedback (Baltes et al., 1989; Yang et al., 2006; Yang et al., 2009). The current findings suggest that older adults do not need any type of external feedback to guide or motivate their retest learning in Stroop interference resolution. Older adults may be self-motivated to optimize and monitor their interference performance. The lack of an individualized and adaptive feedback effect, nevertheless, was surprising.

The linear regression model analyses on the interindividual variability (SDs) in interference RTs suggested that both feedback training groups tended to reduce variances (with a negative slope), though the effect was only significant for the individualized and adaptive feedback group, whereas the no-feedback group did not show this tendency (with a small positive slope). In other words, feedback, particularly individualized and adaptive feedback, had a tendency to gradually reduce the interindividual variability in inhibitory performance. The speculative interpretation for this finding is that with practice, participants may gradually become more skillful at using feedback as an external reminder to help them better regulate their attention to focus on task performance, and, therefore, reduce the chance for outlier performance. The more frequent, trial-by-trial individualized and adaptive feedback may be particularly effective in reducing interindividual interference variability.

Taken together, older adults may not need any external feedback to enhance their retest-induced improvement in average performance to resolve Stroop interference. However, feedback, particularly individualized and adaptive feedback, has a tendency to reduce interindividual interference variability with retest practice.

Limitations and Implications

The current study also has some limitations. First, it does not assess brain activity during training and transfer tasks, and thus provides no information on the possible underlying neural mechanisms responsible for inhibition and its plasticity in the aging brain. Second, lack of transfer effects may be linked to the specific tasks chosen to evaluate transfer effects. We cannot overgeneralize the results to other tasks that may possibly show transfer effects.

Despite these limitations, the current study makes novel contributions to the literature by demonstrating robust plasticity of inhibition among older adults in a well-designed multiple-session training program. These findings suggest that the well-reported inhibitory deficits in older adults are reversible and can possibly reduce through appropriate training or practice. The current study also joins previous work (e.g., Bherer et al., 2005; Dahlin et al., 2008; Kramer et al., 1999) to provide evidence that executive dysfunctions are modifiable among older adults. Going beyond previous work, the current study also suggests that training-induced improvement in inhibition is not item specific and is not likely to be affected by feedback. Older adults may not need any externally provided feedback to direct average performance improvement; however, feedback, especially individualized and adaptive feedback, appears to help reduce interindividual variability in interference effect. These findings hold implications in developing effective self-guided cognitive training programs for older adults.

In summary, three main conclusions can be drawn from the current study: 1) older adults are able to improve their ability to efficiently inhibit irrelevant distracting information with retest practice on a Stroop task across multiple sessions; 2) there are no near or far transfer effects of Stroop inhibitory processing training to certain untrained tasks in older adults; and 3) feedback does not impact the magnitude of inhibitory plasticity with older adults, but it tends to reduce interference variance across retest practice sessions.

References


Thornike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions: III. Functions involving attention, observation and discrimination. Psychological Review, 8, 553–564. doi:10.1037/0039-2128.8.7.553


Received November 8, 2010

Revision received August 17, 2011

Accepted September 1, 2011