



Published in final edited form as:

Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2003 June ; 10(2): 85–98.

Stroop Interference, Practice, and Aging

Douglas J. Davidson¹, Rose T. Zacks², and Carrick C. Williams²

¹University of Illinois at Urbana-Champaign, IL, USA

²Michigan State University, East Lansing, MI, USA

Abstract

We report two experiments that investigate practice effects on Stroop color-word interference in older and younger adults. Both experiments employed a computerized, single-item version of the Stroop task with a voice response, and both involved practice over hundreds of trials. Both experiments showed generally similar practice patterns, including a practice-related reduction in the size of the color-word interference effect. However, the older group continued to show a larger interference effect throughout practice. These findings indicate that older adults show the same trend in practice-related improvement on the Stroop task as younger adults.

As is true for the cognitive literature in general (cf. MacLeod, 1991), the Stroop effect is a mainstay of research on age-related differences in selective attention, automaticity, inhibitory processes, and executive control. A major focus of the aging research has been on the relative size of Stroop interference effects in younger and older adults. The typical finding is that, relative to a baseline condition involving the naming of colors of neutral stimuli (e.g., strings of X's), older adults show a greater increase in reaction time and/or errors in naming of the print colors of incongruent color words than do younger adults (Cohen, Dustman, & Bradford, 1984; Comalli, Wapner, & Werner, 1962; Houx, Jolles, & Vreeling, 1993; Kieley & Hartley, 1997; Li & Bosman, 1996; Spieler, Balota, & Faust, 1996; Vakil, Manovich, Ramati, & Blachstein, 1996; but see Verhaeghen & De Meersman, 1998). Although these data have sometimes been attributed to general slowing effects (Verhaeghen & De Meersman, 1998), others have argued that the larger Stroop effects in older than in younger adults support views proposing age deficits in particular cognitive processes (e.g., the inhibition deficit view of Hasher & Zacks, 1988) or neural mechanisms (e.g., the frontal lobe dysfunction view; Perfect, 1997; Stuss, Eskes, & Foster, 1994; West, 1996). With respect to this last point, there is considerable current interest in relating age differences in the Stroop effect and on other measures of executive function (e.g., task switching; cf. Kramer, Hahn, & Gopher, 1999) to neuroanatomical and neuroimaging findings suggesting that aging particularly affects functions served by prefrontal areas of the brain. For example, a recent fMRI study by Milham et al. (2002) found differences in the patterns of neural activity associated with Stroop performance between younger and older adults, including less extensive activity in the dorsolateral prefrontal cortex in the older group.

MacLeod (1991) stressed that the role of practice is critical for understanding the Stroop interference effect and how the Stroop effect can be used to study selective attention. Theoretical accounts of Stroop performance often link interference from incongruent color words to a presumed greater automaticity of (or familiarity with) word reading as compared to color naming. Prolonged practice of color naming should reverse this discrepancy between color naming and word reading, and the Stroop interference effect should decline in magnitude,

and this has generally been confirmed (Dulaney & Rogers, 1994; Edwards, Brice, Craig, & Penri-Jones, 1996; MacLeod, 1998; MacLeod & Dunbar, 1988; Stroop, 1935). Cognitive aging research could inform this issue, as some have suggested that older adults have difficulty automatizing newly learned skills (Fisk & Rogers, 1991; Kay, 1959; Rogers, 1992). If the practice-related decline in Stroop interference reflects increasing automaticity of color naming and/or the development of an automatic reading suppression response (Dulaney & Rogers, 1994), then the suggestion in the literature that older adults are slower to develop new automatic responses implies that older adults should be less able to develop automaticity in color naming with extended practice. In short, Stroop interference should not decline with practice in the elderly as it does in younger adults, if the development of automaticity is a major factor contributing to the decline in interference, and other factors are held constant. Surprisingly, there has been little systematic study of this question, outside of the Dulaney and Rogers (1994) work.

Dulaney and Rogers (1994) compared older and younger adults in three experiments on a multiple-item version of the Stroop color-word task (28 Stroop words were displayed at a time) and examined practice effects over a series of blocked trials. They found that interference declined with practice for both older and younger adults to approximately the same degree, and also that the pattern of practice-related improvement in interference was similar in both groups (with the greatest reductions occurring early in practice). Nonetheless, despite this similarity in practice effects, performance differences on a post-practice test of reading of the color words from the experiment suggested to the authors that in part, at least, different mechanisms accounted for the reduction in interference in the two age groups. In particular, younger adults were slower to read color words on a post-test compared to a pretest, leading Dulaney and Rogers to argue that the younger adults had developed a word reading suppression response. The older adults did not show a similar slowdown in word reading, and this suggested that word reading suppression did not contribute to the Stroop interference reductions observed with the older adults. For the older adults, and presumably in part for the younger adults, practice-related improvement in Stroop performance was attributed to general task factors such as increased facility of color naming and improvement in scanning strategies.

Although Dulaney and Rogers (1994) obtained consistent results across three experiments, we believe there are a number of reasons to examine practice effects on the Stroop task in the elderly further. First, Dulaney and Rogers used a multiple-item Stroop task rather than a single-item method (i.e., one Stroop word displayed at a time). In addition to being more typical of experimental Stroop studies, single-item Stroop tasks have a number of advantages over multiple-item procedures. In particular, single-item procedures provide performance measures (reaction times and errors) for individual stimuli, thus making it possible to eliminate error trials from the reaction time analyses. Single-item procedures also permit the random intermixing of different trial types (e.g., neutral, interference, facilitation), thus providing for concurrent practice on all conditions, and potentially lessening the influence of set effects. By contrast, all the stimuli in a multiple-item display (whether on a computer monitor or a card) are necessarily presented in the same condition, resulting in intermittent practice on the different conditions and a need to take account of order effects, especially if there are relatively many displays per condition.

Second, it is likely that the multiple-item procedure incurs practice-sensitive processing demands not present in the single-item procedure: As Dulaney and Rogers (1994) note, a multiple-item display requires focusing and systematic scanning processes not needed in single-item conditions. Indeed, there is evidence that practice effects in young adults differ somewhat for single-versus multiple-item Stroop tasks (Edwards et al., 1996; MacLeod, 1998). In particular, by contrast with Dulaney and Rogers' results, MacLeod's (1998) data indicate that young adults do not develop a reading suppression response even after extensive

single-item Stroop color naming practice. If anything, MacLeod found that relative to prepractice reading times, participants read color words more quickly after several hundred trials (Experiment 1) or almost 3000 trials (Experiment 2) of color naming practice rather than more slowly.¹

A further point is that evidence of older adults' greater sensitivity to the visual distraction and extra processing inherent in multiple-item procedures suggests that the differences between the single- and multiple-item procedures may interact with age. For example, age differences on widely used processing speed measures (letter and number comparison, digit-symbol substitution) are larger under the typical multiple-item testing conditions than when only one item is presented at a time (Lustig, Tonev, & Hasher, 2000; Tonev, Lustig, & Hasher, 2000).

From a broader perspective, recent studies in a variety of cognitive domains indicate a considerable interest in the study of age differences in practice effects on cognitive performance. Among other areas, these studies have dealt with visual search (Scialfa, Jenkins, Hamaluk, & Skaloud, 2000), dual-task coordination (Kramer, Larish, Weber, & Bardell, 1999), and task switching (Kramer et al., 1999). Although the findings in each of these studies are of course task-specific, a general picture is beginning to emerge across studies. In particular, it appears that within certain boundary conditions, older adults can benefit at least as much from practice as younger adults. For example, Scialfa et al. (2000) reported similar practice functions in younger and older adults on a consistent-mapping conjunction visual search task as well as evidence indicating equivalent "priority learning" (learning to attend to targets). Similarly, in the task-switching studies carried out by Kramer et al. (1999), older adults initially showed larger task-switching costs than younger adults, but these costs fell more quickly with practice in the older group and after a moderate amount of practice, switching costs were equivalent for the two age groups. Also, the practice benefits were maintained across a 2-month retention interval for both ages. Likewise, Kramer et al. (1999) found that practice under favorable training conditions resulted in reduced and almost equivalent dual-task costs in younger and older adults. Such findings of robust practice effects in older adults, including on tasks that presumably heavily rely on frontal lobe executive functions, are intriguing and set the stage for, among other things, a closer examination of the neural mechanisms involved. Such an analysis has already begun for the Stroop effect. In particular, data from the Milham et al. (2002) fMRI study mentioned earlier indicated that while older adults initially exhibited less activation of parietal areas than young adults in the Stroop task, this pattern changed with practice, so that later in practice, parietal areas were more active than earlier in the task in the older group. Milham et al. interpreted this result as suggesting that Stroop practice benefits in older adults are associated with recruitment of parietal areas. Further investigation of such hypotheses will benefit from detailed behavioral analysis of Stroop practice effects in older adults in a single-item paradigm such as would be used in an event-related fMRI study.

In the light of the above considerations, we carried out two experiments comparing younger and older adults over a large number of trials (approximately 700 in Experiment 1 & 1500 in Experiment 2) in a single-item Stroop task. Neutral and interference trials were randomly intermixed in Experiment 1 as were neutral, interference, and facilitation trials in Experiment 2. Arguably, these conditions better isolate the critical aspects of the Stroop situation (i.e., the potential conflict between the outcomes of color naming and word reading processes) than those used by Dulaney and Rogers (1994). In addition, they provide sufficient data to trace out

¹MacLeod's findings on facilitation effects (faster color naming when the color and the word match relative to color naming of neutral stimuli) may also indirectly speak to the question of whether or not a reading suppression response develops with practice in the single-item Stroop paradigm. In both of MacLeod's experiments facilitation effects were constant across levels of practice. If, as has been argued by MacLeod and Dunbar (1997), Stroop facilitation effects are due to inadvertent reading of the color word, one would have expected facilitation to decline with practice.

in detail how practice affects Stroop performance in younger and older adults, including both general improvements and changes in interference and facilitation.

EXPERIMENT 1

In Experiment 1, participants named the color of color words displayed in a color that mismatched the identity of the word in an interference condition, or strings of Xs displayed in colors in a control condition. These trial types were randomly intermixed. Practice effects were assessed by examining the reductions in average response times and percent error as a function of practice block.

METHOD

Participants—Twenty-four younger adults and 24 older adults from the Michigan State University community participated in Experiment 1 for course credit or monetary compensation (\$10/hr). As shown in Table 1, the older adults had a higher average vocabulary score (Shipley, 1940), out of a possible 40.

Design and Materials—Experiment 1 had a $2 \times 2 \times 6$ mixed design, with Age (young, old) as a between-participants factor; and Stroop Condition (interference, control) and Practice Block (6 consecutive blocks of 128 trials) as within-participant factors. Interference trials were trials in which the color to be ignored mismatched the color to be named (e.g., BLUE displayed in red). Control trials were Xs displayed in the same colors presented in the other trials. The four color words and their corresponding colors on the VGA 16-color palette were blue (#9), green (#10), red (#12), and yellow (#14). The letters of the words and Xs were displayed in uppercase, and the strings of Xs matched the number of characters in the four color words. The four colors were each paired with the three non-matching color words, for a total of 12 distinct interference stimuli which were each presented 48 times (576 interference trials total). The four colors were also paired with the four different strings of Xs to yield 16 distinct control stimuli, each presented 12 times (192 control trials total).

Apparatus and Procedure—The experiment was conducted in a well-lit room using a computer with a 33 cm, 60 Hz color VGA monitor. Stimulus presentation and response measurement were controlled by custom software (Clifton, 1988) in conjunction with a timing card (CyberResearch CYRCTM-05). Voice production latencies were recorded using a Shure microphone placed directly in front of participants (at a distance of approximately 20 cm), along with a Gerbrands G1341T voicekey. Participants were seated in front of the monitor and were given instructions for the Stroop task. They were told that they would be performing a response time task that demanded attention, and that they were to name the colors of stimuli to be displayed on the monitor. They were told that a voicekey connected to a microphone would be registering their response times and that they should speak loud enough to trigger it. A familiarization session consisting of 20 trials of the same type as in the main experiment was also conducted to allow participants to adjust to the task. During this familiarization session, the microphone was positioned and the sensitivity of the voicekey was adjusted to match the loudness of the participant.

In the actual Stroop task, the experimenter pressed a button to begin the experimental task, triggering presentation of the first string. The 768 trials were then presented in 12 blocks of 64 trials, with mandatory 2-min breaks in between blocks. Note that in the data analysis presented below, we combined pairs of these blocks in the analysis to get better estimates of the average response time, resulting in six separate practice blocks. Each trial began with the immediate presentation of a string, which remained on screen until the participant triggered the voicekey. The presentation of the next string began 500 ms after the voicekey was triggered. The trials

continued in this way until the end of a block. The experimenter recorded whether the participant incorrectly responded for each trial.

Data Analysis—Average response times were calculated from interference and control trials from the entire data set, excluding error trials and voicekey failures. All responses were scored as either correct or incorrect based on whether participants named the color of the string presented on screen. Responses that included misnaming, dysfluency, or partial misnaming with recovery were scored as errors. Average error rates were calculated for each condition based on the number of trials each participant saw. Error trials, trials immediately following error trials and trials with RTs less than 200 ms or greater than 5000 ms were excluded prior to the analysis. Using these criteria, 8.9% of the RTs were excluded from the analysis for the younger adults, and 16.3% for the older adults. Most of the excluded trials were error trials and the trials that immediately followed.

Practice effects were analyzed for the interference and control conditions by examining average response times and error rates in six successive blocks of 128 trials. Each block contained a variable number of interference and control trials because of the randomization procedure; each block contained a median of 96 interference trials (range: 88–105), and a median of 32 control trials (range: 24–40). In the analysis of variance results presented below, univariate analyses were corrected using the Huynh–Feldt correction (Everitt, 1995; Huynh & Feldt, 1976; Huynh & Mandeville, 1979) where applicable, to correct for violations of homogeneity of variance assumptions in the repeated measures data.

RESULTS

Response Times—Figure 1A and B show average response time and error rates for interference and control conditions as a function of practice block for older and younger adults. Relative to younger adults, older adults were slower to respond to interference trials compared to control trials. In both age groups, participants were faster with practice, with the greatest improvement in response time appearing in the initial blocks in the interference condition, with less practice-related improvement after the first two blocks.

An analysis of variance of the average response times showed main effects of Age, $F(1, 46) = 22.7, p < .001, MSE = 126670$, due to slower responding of the older adults, Stroop Condition, $F(1, 46) = 230.4, p < .001, MSE = 9229$, indicating slower performance in the interference condition, and an interaction between Age and Stroop Condition, $F(1, 46) = 11.5, p < .001, MSE = 9229$, consistent with a greater Stroop interference effect for the older adults. There was a main effect of Practice Block, $F(5, 230) = 14.2, p < .001$ (H–F epsilon: 0.5755), $MSE = 3103$, and an interaction between Age and Practice Block, $F(5, 230) = 4.8, p = .004$, (H–F epsilon: 0.5755), $MSE = 3103$; suggesting that the older adults improved more with practice than the younger adults. Note that the decrease in average RT with practice in the younger adults was modest at best. Also, there was an interaction of Stroop Condition and Practice Block, $F(5, 230) = 5.43, p < .001$ (H–F epsilon: 0.7852), $MSE = 899$, due to the greater practice-related improvement in the interference conditions compared to the control. The interaction between Age, Stroop Condition and Practice Block was not significant.

The practice effects for the interference and control conditions are well described as a power function of practice block, as Figure 1 shows. The functions plotted in Figure 1 were obtained by fitting each participant's average RT with a two-parameter power function (e.g., $RT = a + \text{Block}^b$) using a linear regression of $\log(\text{RT})$ on $\log(\text{Block})$, and then averaging the resulting parameters over participants. These average parameter values are shown in the equations in Figure 1 (the slope of the regression analysis corresponds to the exponent, which represents the degree of (negative) acceleration in the power function; the intercept corresponds to the multiplier, which functions as a scaling parameter). For the older adults, the obtained fits for

individual participants ranged from poor to good: mean $R^2 = .491$, range = .001 – .951 for the interference condition, and mean $R^2 = .399$, range = .001 to .823, for the control condition. For the younger adults, the fits were similar: mean $R^2 = .362$, range = .001 – .948 for the interference condition, and mean $R^2 = .353$, range = .001 – .895 for the control condition. The power functions in Figure 1 show that participants (in both age groups) improved more in the interference condition relative to the control condition, and that the older adults improved more than the younger adults, although the age difference in improvement did not depend on the Stroop condition involved. Collapsing across the two age groups, in the interference condition the average value for the intercept term was 2.899 (antilog: 794 ms) and the average slope, $-.048$. For the control condition, the average intercept was 2.815 (antilog: 653 ms) and the average slope, $-.026$. The average slope parameters were all significantly less than zero ($p < .05$), based on single-sample t tests. The slope parameters were compared across age groups and Stroop condition in a two-way mixed-model ANOVA, which showed a main effect of Stroop Condition, $F(1, 46) = 10.660$, $p < .005$, $MSE = 0.0001$; and a main effect of Age, $F(1, 46) = 5.728$, $p = .02$, $MSE = 0.006$; but no Age \times Stroop Condition interaction. The main effect of Stroop condition supports the pattern of greater improvement in performance in the interference condition practice shown in Figure 1. The intercept parameters were also entered into a mixed-model ANOVA, and there were main effects of Stroop Condition, $F(1, 46) = 207.456$, $p < .001$, $MSE = 0.001$; and Age, $F(1, 46) = 26.298$, $p < .001$, $MSE = 0.001$; but no Age \times Stroop Condition interaction. The main effects show that older adults were overall slower than the younger adults, and that participants in general were slower in the interference condition.

Some studies adopt a ratio transform to adjust individual average response times for group differences in response speed (e.g., Dulaney & Rogers, 1994; Hartley, 1993; Spieler et al. 1996), and we conducted a similar analysis. We calculated interference ratios by taking a ratio of the average RT for the interference condition minus the control condition to the control condition. This was based on participants' average RTs for each practice block. The interference ratios, like the mean RT analysis presented above, showed that older adults (Ave., .21) showed more interference than the younger adults (Ave., .16), and that the interference ratios declined over practice for both age groups (Block 1 Ave., .22, Block 6 Ave., .17). An analysis of interference ratios showed that there were main effects of Age, $F(1, 46) = 4.91$, $p = .032$, $MSE = 0.038$; and Practice Block, $F(5, 230) = 4.03$, $p = .003$ (H-F epsilon: 0.8390), $MSE = 0.004$; but no interaction between Age and Practice Block.

Error Rates—The average error rates followed much the same pattern as the average response times, as shown in Figure 1. A mixed-model analysis of variance confirmed significant main effects for Age, $F(1, 46) = 17.7$, $p < .001$, $MSE = 0.020$; Stroop Condition, $F(1, 46) = 51.5$, $p < .001$, $MSE = 0.004$, and Practice Block, $F(5, 230) = 18.9$, $p < .001$, $MSE = 0.003$ (H-F epsilon: 0.6134). There was an interaction between Age and Practice Block as well, $F(5, 230) = 5.64$, $p = .001$, $MSE = 0.003$ (H-F epsilon: 0.6134). These results suggest that, as in the response time analysis, performance improved with practice, although perhaps to a greater extent in the older adults. Unlike the response time data, however, there were no interactions between Age and Stroop Condition, or Stroop Condition and Practice Block.

DISCUSSION—Experiment 1 provides evidence that older adults exhibit greater Stroop color-word interference than younger adults. The greater interference, as shown by the analyses of the RT means and interference ratios, is consistent with other studies in the literature showing that older adults exhibit greater interference, not only with color-word interference (Houx et al. 1993; Spieler et al. 1996), but with other forms of Stroop interference as well (Rogers & Fisk, 1991).

Experiment 1 also shows that the effect of practice on the magnitude of Stroop interference (for both response times and error rates) is basically similar in younger and older adults, supporting the original observations of Dulaney and Rogers (1994). There is some evidence in this experiment that the overall effect of practice (i.e., in both control and interference conditions) was larger in the older adults. In both the response times and error rates, it appeared that the older adults gained slightly more in performance in the initial practice blocks. Note, however, that the greater improvement with practice for the older adults did not depend on the Stroop condition involved.

EXPERIMENT 2

The results from the first experiment showed that majority of the practice-related improvement in performance occurred within the first block of practice, and that this first block was associated with more errors than the remaining blocks. Also, the practice effects were assessed over a single session, with fewer trials than are sometimes used in practice studies. It would be useful to confirm that the same pattern holds for longer periods of practice. Thus, the second experiment examined the practice effects over a larger number of blocks. In addition, we added more familiarization trials prior to the start of the practice blocks and screened participants for vision problems. The error rates in Experiment 1 were non-negligible, and this could have been the result of uncertainty over the instructions, or because of difficulty discriminating the presented colors. The additional familiarization trials and screening were intended to reduce these possible sources of error.

With the single-item Stroop task, participants are not able to anticipate the type of Stroop stimulus that will be presented on a given trial, in contrast with the multiple-item display task. Even so, the proportion of the different trial types (e.g., interference or control) can affect how participants perform. MacLeod (1991) proposed that the proportion of different trial types can affect how participants prioritize different dimensions of the Stroop stimuli, for example. MacLeod (1991) also suggested that the effect of adding facilitation trials is to increase the tendency for participants to read the word when the Stroop stimulus is presented, which should lead to poorer performance on interference trials. MacLeod (1991; p. 177) recommended that studies investigating Stroop interference should initially examine performance with interference and control trials (as in Experiment 1), and then later add facilitation trials to determine whether the hypothesized change in priority modifies the pattern of performance observed. Thus, in Experiment 2, we added facilitation trials to the same basic design of Experiment 1 to determine whether the similar pattern of practice-related improvement in the two age groups would hold up when participants have a greater tendency to engage word reading.

The addition of facilitation trials may provide indirect evidence on the development of a reading suppression response that Dulaney and Rogers (1994) suggested developments for young adults practicing a multiple-item Stroop task. As indicated earlier (Footnote 1), if the source of facilitation is the inadvertent reading of the word in a Stroop color-word stimulus, then the development of a reading suppression process should decrease the amount of facilitation shown. If a reading suppression response develops with practice in either age group, that group could show a decrease in facilitation effects at later stages of practice.

In addition, we changed the control condition from Xs to actual words matched for frequency with the color words. Previous investigators have noted that different choices of neutral baseline measures can influence the magnitude of the Stroop effects observed (e.g., Dalrymple-Alford, 1972). If Xs are used, the amount of interference that is calculated by subtracting the control condition from interference may be larger by as much as 65 ms, relative to use of non-associated word distractors as control items (MacLeod, 1991).

METHOD

Participants—Twenty-four younger adults and 24 older adults from the Michigan State University community participated in Experiment 2 for course credit or monetary compensation (\$10/h). As shown in Table 1, the older adults in Experiment 2 had more years of education, on average, and a higher average vocabulary score (Shipley, 1940), out of a possible 40.

In the course of conducting Experiment 1, some of the older participants noted that they had some initial difficulty distinguishing some of the colors presented during the experiment. In Experiment 2, we changed our experimental color yellow to white, and we tested for color vision difficulties as well as acuity prior to running the experiment. Two participants in Experiment 2 did not complete the experiment because of vision problems: 1 older participant did not pass the color vision test, and another older participant reported macular degeneration. These participants were replaced in our sample.

Design and Materials—In addition to the variables examined in Experiment 1, Experiment 2 introduced a facilitation condition. The design was therefore a $2 \times 3 \times 12$ mixed design with Age (young, old), Stroop Condition (interference, control, facilitation), and Practice Block (12 blocks of 128 trials). The practice blocks were provided over two separate sessions (conducted on separate days). The average number of days between the first and second practice sessions was 5.7 days for the older adults and 7.7 days for the younger adults.

In addition to the same types of trials as in Experiment 1, additional facilitation trials were added, consisting of a color word displayed in the same color as named by the color word (e.g., BLUE displayed in blue). Each participant saw 1536 trials in total over the two sessions. Participants saw 880 interference, 380 control, and 276 facilitation trials in all (note that approximately this number of trials were used in the response time analysis, as some trials were errors or voicekey failures – these trials were excluded from analysis of the response times and not replaced). To examine practice effects, we examined 12 consecutive blocks of 128 trials (six blocks each experimental session). Since the colors and color words were randomly intermixed, different numbers of interference, control, or facilitation trials were present in each block for each participant, although the actual number of trials found in each block did not vary a great deal.

The color yellow was changed to white (VGA palette index #15), and the adjectives DEEP, BAD, MAIN, and POOR, displayed in uppercase, were used instead of Xs as the control words. These words have token frequencies of approximately 110, 143, 121, and 113 per million, respectively, by the estimate from Francis and Kucera (1982), and therefore belong in a similar frequency range (relatively high frequency) as blue (165), red (180), green (112), and white (361).

Apparatus and Procedure—The apparatus and procedure for Experiment 2 was the same as in Experiment 1, unless otherwise noted. All participants were tested both in the morning (each session starting between 8:00 a.m. and 9:00 a.m.) and late afternoon (each session starting between 16:00 p.m. and 17:00 p.m.; see Footnote 2). All participants were tested individually.

Prior to receiving instructions for the experimental task, participants were screened with both a color vision test and an acuity test. Participants named two series of colored Xs displayed individually at the center of the computer monitor for the color vision test. All colors presented in the main experimental task were presented during the test, and participants named two consecutive blocks of 10. Participants who could not correctly name the colors were asked to complete the color test again, and those participants who could not correctly name the colors were not run in the main experimental task. The acuity test consisted of a letter naming task

viewed from a distance of approximately 60 cm. Participants were asked to read successive letters of a standardized vision test. No participants had difficulty with the acuity task. Participants performed familiarization trials prior to beginning the main experiment, as in Experiment 1. Importantly, however, there were more familiarization trials (64, compared to 20 in Experiment 1) in the second experiment, and participants completed these 64 familiarization trials at the start of the second session as well. In the main experimental task, participants named colors as in Experiment 1. Participants were given 2-min breaks every 64 trials, and the experiment lasted approximately 1 hr in each session.

Data Analysis—Response times and errors were analyzed as in Experiment 1. Error trials, trials immediately following error trials, and trials with RTs less than 200 ms or greater than 5000 ms were excluded prior to the analysis. Using these criteria, 4.0% were excluded for the young, 4.8% for the old.

RESULTS

Response Times—Figure 2A and B show the average response time and error rates for interference, control, and facilitation conditions as a function of practice block for older and younger adults.² As in Experiment 1, older adults were slower than younger adults in the interference condition compared to the control condition. No age differences in facilitation are apparent. Average response time declined with practice, with the greatest decline in response time appearing in the initial blocks. Unlike Experiment 1, however, error rates remained constant, and did not exhibit a practice-related decline (presumably because they were very low to begin with). Overall, the control, interference, and facilitation condition response times show a similar pattern of practice-related improvement, and this pattern is largely the same in older and younger adults. One difference appears as slight increase in response time at Block seven of the practice blocks for the older adults, but note that Block seven was the start of the second day of practice for this task.

An analysis of the average response times showed main effects of Age, $F(1, 46) = 46.8, p < .001, MSE = 270507$, and Stroop Condition, $F(2, 92) = 225, p < .001, MSE = 8856$ (H-F epsilon: 0.9322); and an interaction between Age and Stroop Condition, $F(2, 92) = 19.6, p < .001, MSE = 8856$ (H-F epsilon: 0.9322). The interaction between Age and Stroop condition was due to greater interference of the older adults relative to the control condition, $F(1, 46) = 22.67, p = .001, MSE = 203745$; but no greater facilitation for older adults relative to the control condition than the younger adults, $F(1, 46) = 2.33, p = .13, MSE = 149426$. There was a main effect of Practice Block as well, $F(11, 506) = 9.3, p < .001, MSE = 5543$ (H-F epsilon: 0.5202), and a marginally significant interaction of Practice Block with Stroop Condition, $F(22, 1012) = 1.626, p = .07, MSE = 1293$ (H-F epsilon: 0.6507). There were no significant interactions of practice with age, however, including no significant interaction of Age, Stroop Condition, and Practice Block.

As in Experiment 1, the practice effects were well described as a power function of practice, as shown in Figure 2. The parameters shown were obtained by fitting each participant's average RT with a two-parameter power function (e.g., $RT = a + \text{Block}^b$) using a linear regression of $\log(RT)$ on $\log(\text{Block})$, and then averaging the resulting parameters over participants, as in Experiment 1. For the older adults, the obtained fits for individual participants ranged from poor to good: mean $R^2 = .369$, range = 0.001–0.865 for the interference condition; and mean $R^2 = .292$, range = 0.001–0.918 for the control condition. For the younger adults, the fits were

²Initially, the design of Experiment 2 included a time of day factor. An analysis of this data revealed only small effects of time of day (e.g., less than 20 ms), and they did not interact with the magnitude of Stroop interference shown by participants (our main concern here). Most importantly, the expected Age \times Time of Day \times Stroop Condition interaction did not materialize. Therefore, the rest of the analysis omits time of day.

similar: mean $R^2 = .395$, range = 0.001–0.818 for the interference condition; and mean $R^2 = .368$, range = 0.001–0.801 for the control condition. The power functions show a slightly greater degree of improvement in the interference condition relative to the control condition, but qualitatively, the practice curves are similar. Collapsing across age, the log–log regression estimates showed that the average slope for the interference condition was $-.041$, the control condition, $-.031$, and the facilitation condition, $-.031$. The average slopes were all significantly less than zero ($ps < .05$), based on single-sample t tests. The average intercept for the interference condition was 2.901 (antilog: 796 ms), the control condition 2.846 (antilog: 702 ms), and the facilitation condition 2.823 (antilog: 665 ms).

The slope parameters were compared across age groups and Stroop condition in a two-way mixed-model ANOVA, which showed a marginally significant main effect of Stroop Condition, $F(2, 92) = 2.023$, $p = .053$, $MSE = 0.0007$; but no main effect of Age, nor Age \times Stroop Condition interaction. Two planned comparisons revealed that the Stroop Condition main effect was due to a difference between the interference and control conditions, $F(1, 47) = 6.503$, $p = .014$, $MSE = 0.0009$; but no corresponding difference between facilitation and control conditions. Thus, this analysis supports the suggestion that participants improved with practice to a greater degree in the interference condition compared to the control and facilitation conditions, and importantly, that there were no age differences in the rate of improvement for the different conditions. The intercept parameters were also entered into a two-way mixed-model ANOVA. It indicated main effects for Stroop Condition, $F(2, 92) = 112.6$, $p = .001$, $MSE = 0.0036$; and Age, $F(1, 46) = 37.25$, $p < .01$, $MSE = 0.0549$. The interaction between Stroop Condition and Age was also significant, $F(2, 92) = 3.256$, $p = .001$, $MSE = 0.0036$. Planned comparisons between interference and control conditions were significant, $F(1, 46) = 128.1$, $p < .001$, $MSE = 0.0062$; as well as between facilitation and control conditions, $F(1, 46) = 23.9$, $p < .001$, $MSE = 0.0052$. These comparisons were also conducted with Age as a factor, revealing a marginally significant interaction with Age comparing the interference condition to the control condition, $F(1, 46) = 3.704$, $p = .06$, $MSE = 0.0062$, but no interaction with Age comparing facilitation to control. The intercept parameter analysis thus supports the main analysis in showing the basic interference and facilitation effects, as well as providing some support for the greater effect of interference in the older adults.

To determine whether the same pattern of results holds for a ratio measure of interference or facilitation, two separate mixed ANOVAs were performed using participants' average interference ratios and facilitation ratios respectively, entering the same factors as the average response time. The ratios showed much the same pattern as the average response time measures. For the interference ratios, this analysis indicated main effects of Age: $F(1, 44) = 11.5$, $p < .001$, $MSE = 0.031$, showing that older adults had a higher interference ratio (*Ave.* 0.143) than the younger adults (0.094). The analysis of the facilitation ratios did not reveal any statistically significant differences.

Error Rates—The error rates were lower in the second experiment, and also unlike Experiment 1, no age differences in overall error rate were found, nor was there an Age \times Stroop Condition interaction, as indicated in the pattern in Figure 2. There was a main effect of Stroop Condition: $F(2, 88) = 80.67$, $p < .001$, ($H-F$ epsilon: 0.7478), $MSE = 0.002$. No effects due to practice were observed. Two contrasts showed that the main effect of Stroop condition reflected a higher average error rate in the interference relative to the control condition: $F(4, 44) = 21.07$, $p < .001$, $MSE = 0.005$; and a higher average error rate in the control condition relative to the facilitation condition, $F(4, 44) = 5.002$, $p < .005$, $MSE = 0.015$.

DISCUSSION—Experiment 2 produced much the same pattern as Experiment 1 with respect to the Stroop interference effect and practice. These results, as in Experiment 1, replicate many previous findings in the literature suggesting that older adults exhibit greater Stroop color-word

interference (cf. Verhaeghen & De Meersman, 1998). Also, the lack of an age by practice block interaction with Stroop condition demonstrates a similar pattern of age-related improvement in performance as demonstrated with the multiple-item version of the task (Dulaney & Rogers, 1994). This result was obtained when facilitation trials were added to the design, as well as a different baseline condition, suggesting that the similar pattern of practice-related improvement in older and younger adults is robust under conditions in which participants are more likely to read the distracting word.

One potential difference between Experiments 1 and 2 with respect to practice effects was the finding of an interaction between age and practice block observed in the average response times and error rates in Experiment 1, but not Experiment 2. Thus, it appeared in Experiment 1 that older adults improved slightly more with practice (no more so for the interference condition, however). When more practice trials were provided in Experiment 2, however, there was no evidence of an age difference in the rate of performance improvement. This suggests that participants were more familiar with the task in Experiment 2 than Experiment 1, and this could be the reason for the difference in results. On the balance, the results of both experiments provide good evidence of similar patterns of practice-related reduction of Stroop interference in old and young adults.

In addition, Experiment 2 provided little evidence that the rate of improvement in the facilitation condition was different from the baseline condition in the younger adults. A difference would be expected if younger adults were developing a word reading suppression process during practice. Also, the pattern of improvement for the facilitation condition compared to the control condition appeared to be the same for the two age groups. If younger adults, but not older adults were developing a word-reading suppression process, then an interaction with age would have been expected for the practice effect differences between baseline and facilitation conditions. A potential caveat to this finding, however, is that the facilitation effects observed in Experiment 2 were relatively small, and therefore potentially less sensitive to practice related improvement.

GENERAL DISCUSSION—Extended practice on the Stroop color-word task appears to result in a complex of effects. Overall response time improves in both control and interference conditions. There is a reduction in the size of the Stroop interference effect indexed by the difference between interference and control conditions. These results appear to generalize across different versions of the Stroop task, as shown here by the different baseline conditions and the addition of the facilitation condition in Experiment 2, and by the basic similarity in pattern with previous investigations by MacLeod (1998) and others. Also, as was found by MacLeod, when the facilitation condition is included, there is little change in the size of the difference between facilitation and control conditions with practice, supporting MacLeod's contention that different mechanisms underlie Stroop facilitation and interference effects. Further, and most importantly for present purposes, there are no age differences in any of the above patterns.

Thus, the present findings replicate other results in the literature (e.g., Dulaney & Rogers, 1994), showing similar patterns of practice-related improvement in Stroop interference for older and younger adults. This suggests that older adults can improve performance in both multiple- and single-item versions of the Stroop task, even though there appear to be important differences between the two paradigms (as outlined in the introduction). Another point of similarity between our data and those Dulaney and Rogers is that even at the end of practice, the Stroop effect was larger in the older group. One potential difference between our Experiments 1 and 2 with respect to practice effects was the finding of an interaction between age and practice block observed in the average response times and error rates in Experiment 1, but not Experiment 2. Thus, it appeared in Experiment 1 that older adults improved slightly

more with practice (no more so for the interference condition, however). When more practice trials were provided in Experiment 2, however, there was no evidence of an age difference in the rate of performance improvement. This suggests that participants were more familiar with the task in Experiment 2 than Experiment 1, and this could be the reason for the difference in results. On the balance, the present results provide good evidence of similar patterns of practice-related reduction of Stroop interference in old and young adults.

The current results are also similar to the findings (mentioned above) from studies of practice effects in visual search (Scialfa et al., 2000), dual-task coordination (Kramer et al., 1999), and task switching (Kramer et al., 1999). In all these cases, the data indicate robust practice effects in older adults that are at least equivalent to those of younger adults under many (but likely not all) circumstances. Furthermore, both of the Kramer et al. studies demonstrated good retention of practice-related improvement over retention periods of up to 2 months in older as well as younger adults. Likewise in our data, the only indication of a loss of practice benefits over the average 5.7 days between sessions in Experiment 2 was the older adults' slight increase in reaction times for the first block of session 2.

Despite this evidence of similarity of age patterns in practice effects across different tasks and across different versions of the Stroop task, we can make only tentative suggestions about the mechanisms of practice effects and their relationship to age. With respect to the Stroop effect, it is likely that several factors are involved, including non-specific performance effects of practice (e.g., stimulus encoding, response execution, & color name facility) that impact both control as well as interference conditions. The involvement of a reading suppression response, as was reported by Dulaney and Rogers (1994) for younger but not older adults, cannot be ruled out by our data, given that we did not collect post-test color word reading times. Nonetheless, taken together, a number of studies at least suggest that this was not the central factor underlying the practice effects for either age group in our data. These considerations include MacLeod's (1998) failure to find evidence of the development of a reading suppression response on post-practice color word reading measures, the constancy of facilitation effects over different levels of practice in both MacLeod's and our results, and the absence of age differences in practice effects in our data.

ACKNOWLEDGEMENTS

A portion of the research reported here was presented at the 1998 meeting of the Cognitive Aging Conference in Atlanta, GA. This research was supported by a National Institute on Aging grant AGO 4306 awarded to Lynn Hasher and Rose Zacks.

REFERENCES

- Clifton C. PC-Experiment [Computer File]. 1988
- Cohen NC, Dustman RE, Bradford DC. Age-related decrements in Stroop color test performance. *Journal of Clinical Psychology* 1984;40:1244–1250. [PubMed: 6490922]
- Comalli PE, Wapner S, Werner H. Interference effects in childhood, adulthood, and aging. *Journal of Genetic Psychology* 1962;100:47–53. [PubMed: 13880724]
- CyberResearch Inc. Branford, CT, USA:
- Dalrymple-Alford EC. Associative facilitation and interference in a color-naming task. *Perception and Psychophysics* 1972;28:209–210.
- Dulaney CL, Rogers WA. Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1994;20:470–484.
- Edwards S, Brice C, Craig C, Penri-Jones R. Effects of caffeine, practice, and mode of presentation on Stroop task performance. *Pharmacology, Biochemistry, and Behavior* 1996;54:309–315.

- Everitt BS. The analysis of repeated measures: A practical review with examples. *The Statistician* 1995;44:113–135.
- Francis, WN.; Kucera, H. *Frequency analysis of English usage: Lexicon and grammar*. Houghton Mifflin; Boston, MA: 1982.
- Hartley AA. Evidence for the selective preservation of spatial selective attention in old age. *Psychology and Aging* 1993;3:371–379. [PubMed: 8216957]
- Hasher, L.; Zacks, RT. Working memory, comprehension, and aging: A review and a new view. In: Bower, GH., editor. *The psychology of learning and motivation*. 22. Academic Press; San Diego, CA: 1988. p. 193-225.
- Houx PJ, Jolles J, Vreeling FW. Stroop interference: Aging effects assessed with the Stroop Color-Word Test. *Experimental Aging Research* 1993;19:209–224. [PubMed: 8223823]
- Huynh H, Feldt LS. Estimation of the Box correction for degrees of freedom for sample data in randomized block and split-plot designs. *Journal of Educational Statistics* 1976;1:69–82.
- Huynh H, Mandeville GK. Performance of traditional *F* tests in repeated measures designs under covariance heterogeneity. *Communications in Statistics – Theory and Methods* 1979;A9:61–74.
- Kay, H. Theories of learning and aging. In: Birren, JE., editor. *Handbook of aging and the individual*. University of Chicago Press; Chicago: 1959. p. 614-654.
- Kieley JM, Hartley AA. Age-related equivalence of identity suppression in the Stroop color-word task. *Psychology and Aging* 1997;12:22–29. [PubMed: 9100265]
- Kramer AF, Hahn S, Gopher D. Task coordination and aging: Explorations of executive control processes in the task switching paradigm. *Acta Psychologica* 1999;101:339–378. [PubMed: 10344190]
- Kramer, AF.; Larish, JL.; Weber, TA.; Bardell, L. Training for executive control: Task coordination strategies and aging. In: Gopher, D.; Koriat, A., editors. *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*. MIT Press; Cambridge, MA: 1999. p. 617-652.
- Li KZ, Bosman EA. Age differences in Stroop-like interference as a function of semantic relatedness. *Aging, Neuropsychology, and Cognition* 1996;3:272–284.
- Lustig, C.; Tonev, S.; Hasher, L. Visual distraction and processing speed I; Poster presented at the 2000 Cognitive Aging Conference; Atlanta, Georgia. Apr. 2000
- MacLeod CM. Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin* 1991;109:163–203. [PubMed: 2034749]
- MacLeod CM. Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory and Cognition* 1998;26:201–211.
- MacLeod CM, Dunbar K. Training and Stroop-like interference: Evidence or a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1988;14:126–135.
- Milham MP, Erickson KI, Banich MT, Kramer AF, Webb A, Wszalek T, Cohen NJ. Attentional control in the aging brain: Insights from an fMRI study of the Stroop task. *Brain and Cognition* 2002;49:277–296. [PubMed: 12139955]
- Perfect, T. Memory aging as frontal lobe dysfunction. In: Conway, MA., editor. *Cognitive models of memory*. Psychology Press; Sussex: 1997. p. 315-339.
- Rogers WA. Age differences in visual search: Target and distractor learning. *Psychology and Aging* 1992;7:526–535. [PubMed: 1466821]
- Rogers WA, Fisk AD. Age-related differences in the maintenance and modification of automatic processes: Arithmetic Stroop interference. *Human Factors* 1991;33:45–56. [PubMed: 2037308]
- Scialfa CT, Jenkins L, Hamaluk E, Skaloud P. Aging and the development of automatic conjunction search. *Journal of Gerontology: Psychological Sciences* 2000;55:P27–P46.
- Shipley WC. A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology* 1940;9:371–377.
- Spieler DH, Balota DA, Faust ME. Stroop performance in normal older adults and individuals with Senile Dementia of the Alzheimer's Type. *Journal of Experimental Psychology: Human Perception and Performance* 1996;22:461–479. [PubMed: 8934854]

- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 1935;18:643–662.
- Stuss, DT.; Eskes, JK.; Foster, JK. Experimental neuropsychological studies of frontal lobe functions. In: Boller, F.; Grafman, J., editors. *Handbook of neuropsychology*. 9. Elsevier Press; Amsterdam: 1994.
- Tonev, S.; Lustig, C.; Hasher, L. Visual Distraction and Processing Speed II; Poster presented at the 2000 Cognitive Aging Conference; Atlanta, Georgia. Apr. 2000
- Vakil E, Manovich R, Ramati E, Blachstein H. The Stroop color-word test as a measure of selective attention: Efficiency in the elderly. *Developmental Neuropsychology* 1996;12:313–325.
- Verhaeghen P, De Meersman L. Aging and the Stroop effect: A meta-analysis. *Psychology and Aging* 1998;13:120–126. [PubMed: 9533194]
- West RL. An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin* 1996;120:272–292. [PubMed: 8831298]
- Zacks, RT.; Hasher, L. Directed ignoring: Inhibitory regulation of working memory. In: Dagenbach, D.; Carr, T., editors. *Inhibitory mechanisms in attention, memory, and language*. Academic Press; New York: 1994. p. 241-264.

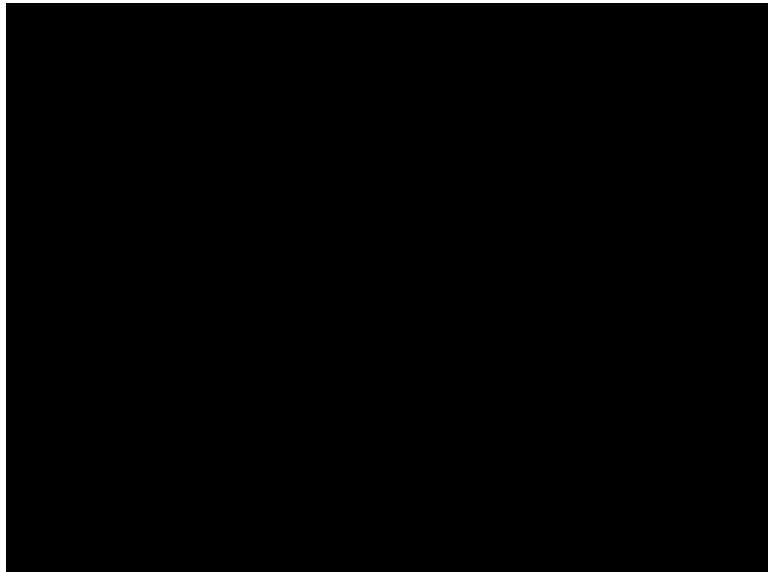


Fig. 1.
Average Response Times and Error Rates by Practice Block, Experiment 1.

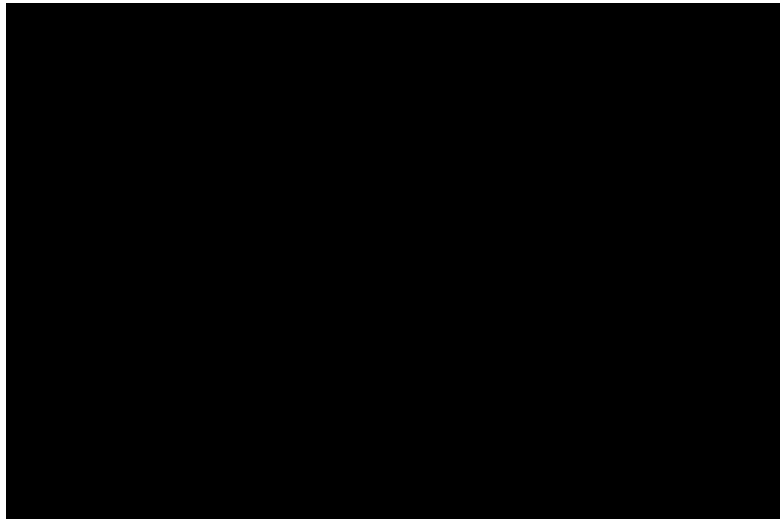


Fig. 2.
Average Response Times and Error Rates by Practice Block, Experiment 2.

Table 1
Participant Data for Experiments 1 and 2 (Standard Deviation in Parentheses).

	Young (<i>n</i> = 24)	Old (<i>n</i> = 24)
Experiment 1		
Age (years)	20.6 (3.5)	73.4 (5.0)
Education ^a	14.5 (1.3)	15.4 (2.7)
Vocabulary ^b	30.8 (3.6)	35.3 (4.6)
Experiment 2		
Age (years)	20.3 (2.8)	74.9 (4.5)
Education ^a	14.0 (1.6)	15.6 (2.6)
Vocabulary ^b	30.0 (3.0)	35.0 (3.0)

Note.

^a Years of formal education; Experiment 1: $t(46) = 1.43$, *ns*; Experiment 2: $t(46) = 2.52$, $p < .01$.

^b Average score, number correct out of 40; Experiment 1: $t(46) = 3.81$, $p < .001$; Experiment 2: $t(44) = 5.70$, $p < .01$.