

Training on integrated versus separated Stroop tasks: The progression of interference and facilitation

COLIN M. MACLEOD

University of Toronto, Scarborough, Ontario, Canada

Two experiments examined the course of interference and facilitation in the Stroop (1935) task during training. Two versions of the task were compared: integrated (e.g., the word RED in the color green) and separated (e.g., green asterisks above the word RED). Stimuli were congruent (RED in red), incongruent (GREEN in red), or neutral (xxx in red). Over 5-day (Experiment 1) and 10-day (Experiment 2) training sessions, facilitation due to congruence was small, stable, and equivalent in both task versions. In contrast, interference declined sharply on the integrated task over Days 1–3, then slowed to parallel the gradual decline on the separated task. Finally, training on the color naming task did not affect a word reading task administered after training. These findings imply that (a) Stroop interference initially reflects two problems—overcoming integration and managing two conflicting information sources; (b) with practice, the larger integration problem is solved relatively quickly, rendering the integrated and separated tasks quite comparable thereafter; and (c) facilitation and interference in the Stroop task may be independent. These results challenge extant theories of the Stroop effect, which do not predict such effects.

Perhaps uniquely among the well-known phenomena in cognition, research on the Stroop (1935) effect continues to increase decades after the task was introduced. This is evident in citation data (see MacLeod, 1992) and in the publication dates for articles cited in the most recent literature review (MacLeod, 1991). Each year, more articles explore this phenomenon, still one of the most intriguing in all of psychology, even after 60 years.

Theoretical Background and Research Goals

Why is this phenomenon so compelling? Part of the reason lies in its size and ease of demonstration. One can literally feel the interference from the incongruent word when trying to name the color in which it is printed (e.g., to say “red,” not “green,” to the word GREEN printed in red ink). But it is not just the empirical power of the Stroop effect that supports its prevalence. This task provides a theoretical window on how we deal with conflicting stimuli and task demands, and it is a fertile testing ground for ideas about automaticity and the role of learning in the development of that automaticity. These are fundamental questions about how attention works (see, e.g., Shiffrin, 1988).

MacLeod's (1991) review article singled out three issues as crucial to understanding the cause(s) of the Stroop

effect and how it speaks to the operation of attention. The first was practice, or *training*. To what extent does the amount of experience a person has with each of the two dimensions—color and word, in the case of the Stroop effect—modulate the interference that will be observed? If automaticity is a continuum (e.g., Logan, 1985; MacLeod & Dunbar, 1988; Shiffrin, 1988), then it is important to understand how automaticity changes with practice. The second issue highlighted was *integration*. To what extent does the degree of melding of the two dimensions influence the extent of interference? This relates directly to how we use attention to select among dimensions. The third issue identified was the relation between facilitation and interference. Does the decreased response time on congruent trials (RED in red, say “red”) stem from the same mechanism(s) as the increased response time on incongruent trials? Some existing work (as reviewed by MacLeod, 1991) suggests that facilitation and interference may not have a common base, contrary to virtually all existing theories. The aim of the present article is to address these three questions simultaneously.

The Question of Training

Training has been of concern in the Stroop task from the beginning (and even in its precursors; see Cattell, 1886), given the imbalance in experience with word reading and color naming. In his seldom-cited Experiment 3, Stroop (1935) himself investigated practice effects. Each day, he had subjects color name all of the words on four cards, each of which contained 50 incongruent color–word combinations (200 trials). Over 8 days, Stroop observed a decline in the interference subjects suffered, from about 418 msec per word on the 1st day to about 82 msec per word

This research was supported by Natural Sciences and Engineering Research Council of Canada Grant A7459. I am grateful to P. J. Lootsteen and Marina Vanayan for assistance with collecting the data of Experiment 1 and to Shelley Hodder for assistance with collecting the data of Experiment 2. I also very much appreciate the comments of Geoffrey Loftus, Arthur Samuels, and an anonymous reviewer on a prior draft. Correspondence should be addressed to C. M. MacLeod, Division of Life Sciences, University of Toronto, Scarborough, ON M1C 1A4, Canada (e-mail: macleod@scar.utoronto.ca).

on the last day. Although this 80% drop is probably a considerable overestimate due to Stroop's design,¹ it is important to realize that interference from incompatible words was still very much present in the color naming task even after all of this practice. Other studies (e.g., Ackerman & Schneider, 1984; Ellis & Dulaney, 1991) have confirmed that although interference decreases with practice, it is very resistant to eradication.

Unlike most researchers, Stroop (1935) investigated both color naming (Experiments 2 and 3) and word reading (Experiments 1 and 3), thereby providing a more complete picture. When asked to read words printed in incompatible colors before the color naming training, subjects in his Experiment 3 showed almost no interference; after training, they showed reliable "reverse Stroop" interference of about 314 msec. This exceptionally large effect held only for four stimulus cards, and then subjects reverted to virtually no interference in word reading, but it still was a provocative finding. Perhaps the extensive practice in color naming had shifted the priority of the color dimension enough that color processing was briefly more automatic than word processing.

Using single items on each trial, MacLeod and Dunbar (1988, Experiment 3) reported evidence that seems consistent with this hypothesis. Subjects had to learn to call four unfamiliar shapes by four consistent color names (e.g., polygon A was called "red," polygon B "yellow," and so on). Prior to practice at shape naming, when the shapes were presented in incompatible colors (e.g., polygon A in yellow), subjects showed interference in naming the shapes but not in naming the colors. Initially, then, color naming was more automatic than shape naming. After 5 days of shape naming training, interference in shape naming and in color naming was roughly symmetrical; after 20 days, only color naming showed interference, reversing the pattern from the 1st day. As training (and hence automaticity) increased on the shape dimension, the direction of interference shifted.

Recently, in a study of cognitive aging, Dulaney and Rogers (1994) provided further confirmation of Stroop's (1935) training patterns. Each of their displays involved multiple items, now 28 at once on a computer screen. For the most direct comparison, I will focus only on the data from their young subjects in Experiment 2. In a single session, subjects did the color naming task on 80 displays each of control and incongruent stimuli. Interference in color naming declined from 340 to 180 msec per item, a 47% decline. Before and after the color naming practice, subjects also did three displays of word reading. Whereas before practice, there was almost no interference (16 msec) for reading incongruent words as opposed to neutral words, after practice, "reverse Stroop" interference was a reliable 74 msec. As in Stroop's experiment, then, Dulaney and Rogers found that practice in color naming led to the emergence of interference in word reading.²

In related tasks, similar patterns appear. As an illustration, Simon, Craft, and Webster (1973) had subjects make

left-right keypress responses to tones of high versus low pitch. The tones were presented to either the left or the right ear, but ear of presentation was irrelevant. Nevertheless, responses were 60 msec faster when the ear and response hand were compatible. Simon et al. then trained their subjects for 5 days. Although practice reduced the Simon effect to 35 msec, it was still clearly present, having declined 42%. Shor, Hatch, Hudson, Landrigan, and Shaffer (1972) reported an analogous pattern over 30 days with a different sort of spatial analogue of the Stroop task.

The two experiments to be presented here extended these prior studies in two main ways. First, the present experiments used the single-item-per-trial procedure, more common in contemporary studies of the Stroop effect. Second, in addition to incongruent and control trials, congruent trials were included to permit examination of facilitation as well as interference. The principal goal was to explore the time course of the Stroop effect over practice at color naming, with the additional aim of determining how that practice in turn affected performance on the word reading task.

The Question of Integration

Traditionally, the Stroop task used words printed in conflicting colors, integrating the two dimensions. But since Dyer's work (Dyer, 1971, 1973; Dyer & Severance, 1973) and with the advent of computerized testing, it has become common to present the word and color separately. This technique is particularly useful for investigation of such manipulations as stimulus onset asynchrony (see, e.g., M. Glaser & W. R. Glaser, 1982) or multiple words in a single display (see, e.g., Kahneman & Chajczyk, 1983; MacLeod & Hodder, 1998; Yee & Hunt, 1991), as well as for variations on the Stroop task, such as the picture-word interference task (see, e.g., Lupker, 1979). But how comparable is the separated version of the task to the integrated version? Do manipulations influence the two types of stimuli correspondingly? Are the same processes involved in dealing with these two types of stimuli?

It is clear that integration versus separation is critical from the landmark work of Garner (1962, 1974) and studies that have followed in that tradition (e.g., Cheng & Pachella, 1984; Dunn, 1983; Monahan & Lockhead, 1977). Yet despite occasional mention (e.g., Gardner & Long, 1962), there has been little work done with regard to integration/separation in the Stroop paradigm. Gatti and Egeth (1978) increased the distance between the color and the word from 1° to 5° and showed that interference declined from 90 to 40 msec. They did not, however, include a direct comparison to an integrated stimulus. In related work, Kahneman and Henik (1981) showed that location interacted with attention: If an incompatible word and color were both at the same attended location, there were 202 msec of color naming interference; if only one was attended, interference fell to 50 msec.

In the most relevant study, Flowers and Stoup (1977) initially used a card sorting version of the Stroop task. They

manipulated both training and whether the word and color were integrated (word printed in color; Experiment 2) or separated (word inside a color rectangle; Experiment 3). Subjects sorted decks of 30 cards by ink color. For integrated stimuli, despite falling considerably, interference was still quite evident after 4 days of practice. In sharp contrast, interference was less to begin with and disappeared completely within 2 days for separated stimuli. But when they switched to a more standard color naming task in their Experiment 4, interference for separated stimuli showed very little decline with practice, and they concluded that the color naming and card sorting tasks behaved differently, with interference in color naming being relatively impervious to training even for separated stimuli. Their practice was relatively minimal, though, and they did not explore what practice in color naming would do to integrated stimuli, nor whether facilitation would respond similarly to interference.

A key aim here was to examine training effects not only on interference but also on facilitation in what will henceforth be called the integrated and separated versions of the Stroop task. At least a couple of predictions can be tested. First, on the basis of the studies just described, and taking into account Garner's (1962, 1974) ideas, it might reasonably be expected that both interference and facilitation would be reduced more in the separated task than in the integrated task. Second, changes with practice should be easier to implement and quicker to appear in the separated version than in the integrated one. These predictions will be examined in the two experiments.

The Question of the Relation Between Interference and Facilitation

The third major concern is the relation between interference in the incongruent condition and facilitation in the congruent condition. As set out in detail by MacLeod and MacDonald (1997), the principal models in the literature all make the claim that interference and facilitation stem from the same mechanism. This is true for Logan's (1980) automaticity model, for the Cohen, Dunbar, and McClelland (1990) and Phaf, van der Heijden, and Hudson (1990) connectionist models, for the W. R. Glaser and M. Glaser (1989) network model, and for Lindsay and Jacoby's (1994) process-dissociation-based model. Yet there is already mounting evidence that conflicts with their shared claim. As the proportion of congruent trials increased, Lowe and Mitterer (1982) found that interference increased monotonically, yet facilitation was unaffected. As they varied stimulus onset asynchrony between color and word through a quite large range, M. Glaser and W. R. Glaser (1982) found different time courses for interference and facilitation. In my laboratory, drastic changes in stimulus characteristics (Dunbar & MacLeod, 1984) and extensive variations in training (MacLeod & Dunbar, 1988) produced very little change in facilitation despite huge effects on interference in the same study. In a large individual differences study, Vanayan (1992) showed that interference and facil-

itation scores were uncorrelated. In a study varying the proportion of color-word trials, Tzelgov, Henik, and Berger (1992) observed that interference increased as the proportion of color-word stimuli increased, but facilitation remained unchanged. Although no single finding is definitive, taken together these results do not fit well with the common mechanism hypothesis for interference and facilitation.

Most recently, MacLeod and MacDonald (1997) have argued that the apparently faster performance on congruent trials is due to inadvertent reading of color words, not true facilitation of color naming. This is in contrast to the true interference that words exert on colors during incongruent trials. In Experiment 1, bilinguals showed facilitation only within one language, not between languages, despite showing interference in both situations. The same dissociation was evident for color-related words such as *lemon* or *blood* (Experiment 2) and for newly learned responses such as saying "pen" to the color red (Experiment 3). Only when accidental reading of the word produced the same response as correct color naming was "facilitation" observed. MacLeod and MacDonald argued that this apparent facilitation was actually due to the mixture of unobservable fast word-reading times into the distribution of slower color naming times. Experiment 4 showed that increasing the probability of reading errors increased apparent facilitation, consistent with these reading errors actually being the cause of facilitation.

In the present work, both facilitation and interference were measured while the fundamental variables of training and integration were manipulated. We know that interference should decline with practice (Ackerman & Schneider, 1984; Ellis & Dulaney, 1991; Stroop, 1935), and we can deduce that it should be smaller for separated than for integrated stimuli (Gatti & Egeth, 1978; Kahneman & Henik, 1981), although direct comparisons have not been made previously. But these effects need not have an impact on facilitation in the same way if facilitation is the result of infrequent inadvertent word reading.

The Logic of the Experiments

The strategy for these experiments was to train subjects on one of two versions of the color naming task—integrated or separated—over a substantial number of days and trials. In Experiment 1, training consisted of 5 days with 108 trials daily; in Experiment 2, it was made up of 10 days with 288 trials daily. Incongruent, control, and congruent conditions were all included to evaluate the effects of practice on both interference and facilitation in color naming. In addition, before and after the color naming training, a measure of word reading was taken to permit exploration of the effects that training on one task has on interference and facilitation in the other task. The first experiment, with more modest training, can be seen as a "minimalist" version of the second one. The goal was to compare practice effects on the integrated and separated versions of the Stroop task directly, using a single-item computerized procedure. The second experiment increased

Table 1
Experiment 1, Word Reading: Mean Facilitation, Interference, and Control Response Times (RT, in Milliseconds) and Corresponding Mean Proportions of Errors (PE) as a Function of Day of Training for Both the Integrated and the Separated Tasks

Condition	Task			
	Integrated		Separated	
	RT	PE	RT	PE
Day 1				
Facilitation	-8	.007	-13	.000
Interference	-12	.000	-6	.000
Control	515	.000	424	.000
Day 5				
Facilitation	13	.000	-11	.000
Interference	24	.000	5	.000
Control	483	.000	418	.000

Note—The control condition values are actual means. The facilitation and interference values are deviations from the corresponding control condition, in which a negative value should accompany facilitation (i.e., faster than control) and a positive value should accompany interference (i.e., slower than control).

training and sample size, as well as altering the apparatus and the stimulus display characteristics, to increase power and generalizability.

EXPERIMENT 1

Method

Subjects. Eight University of Toronto at Scarborough undergraduate students were paid \$24 each to take part. Four subjects were assigned to each task (integrated and separated).

Apparatus and experimental control. Testing was carried out on an Apple IIe computer equipped with an Electrohome EC-1302 RGB color monitor and Supercolor RGB videocard. Timing was synchronized with the screen refresh rate. When the subject spoke into a microphone, a Lafayette voice key connected to the game paddle port was tripped. Programming was done in Applesoft BASIC with machine language millisecond accuracy timing routines accessing a Mountain Computer real time clock. The screen background color was black; the instructions and control stimuli were presented in white. The four colors used were pink, blue, green, and yellow. All information was presented in standard Apple 40-column uppercase format, centered horizontally near the bottom of the screen. The rest of the screen was covered with a cardboard mask, leaving a small window for the stimuli.

Tasks. For the integrated group, the word appeared in standard uppercase font in a particular color on line 19 of the screen. For the separated group, a row of eight same-colored asterisks appeared on line 19 centered directly above the uppercase word printed in neutral white on line 21.

Procedure. On the 1st day, subjects were given general instructions and encouraged to respond as quickly as possible, while simultaneously minimizing errors. They were urged to try to improve every day, and not to let their motivation sag.

The first task was word reading. Subjects were to read each word aloud into the microphone as quickly as possible, ignoring its print color. Each trial began with a 250-msec warning (a row of four asterisks) followed by a 250-msec blank period. Then the color-word stimulus appeared and stayed on until the subject responded. A 250-msec blank screen followed stimulus offset and then a message appeared indicating the correct response (e.g., "RED was correct"), after

which the subject pressed either the "/" key to indicate a correct response or the "z" key to indicate an error of any kind. Any mistake, whether an incorrect response or a response that was not loud enough or the like, was to be designated as an error. Thus, the error data consist of true response errors, vocal stumbles, and equipment problems. The next trial followed 250 msec after the subject indicated his/her accuracy.

Subjects first performed 20 practice trials made up of four non-color words—LION, BEAR, TIGER, and RABBIT. Each word was presented once in each color and once in white (the control condition). The order of presentation of these 20 trials was random. The experimenter was present for the entire practice session in case the subject had any questions or problems following instructions.

During the experimental trials, the subject was left alone and scored the accuracy of his/her own responses.³ There were 108 trials in a session, made up of 36 congruent, 36 incongruent, and 36 control trials. The words were PINK, BLUE, GREEN, and YELLOW. Each word was presented nine times in its own color (the 36 congruent trials), nine times in white (the 36 control trials), and three times in each of the three other colors (the 36 incongruent trials). Order of trials was completely randomized with the constraint that none of the three conditions could occur more than three times in succession. Short subject-controlled breaks were allowed after 36 trials and after 72 trials.

Once the word reading experimental trials were completed, the color naming training trials began. Subjects were told to name the color of print aloud, ignoring the word itself. They again did 20 practice trials and 108 experimental trials. The only change (other than instructions) was that the control condition was now the stimulus XXXX printed in an ink color. After finishing the practice and experimental color naming trials, subjects were asked to return at about the same time the next day.

On Days 2–4, subjects did the 20 practice and 108 experimental color naming trials each day, scoring their own accuracy. The experimenter reminded subjects prior to each session to try to respond as quickly as possible without mistakes. On Day 5, subjects did one last color naming session and then returned to word reading, conducted as on Day 1. Subjects received no feedback at any time other than during practice on the 1st day. They were paid and debriefed at the end of Day 5.

Results and Discussion

For each subject, two critical difference scores were computed: facilitation (congruent minus control) and interference (incongruent minus control). Thus, interference scores should be positive and facilitation scores should be negative. All analyses of variance (ANOVAs) were conducted both on the raw mean response times (RTs) and on these difference scores, but because interpretation emphasizes interference and facilitation, only those on the difference scores are reported. The same conclusions are supported by both sets of analyses. Furthermore, the relevant control condition means are shown with each analysis to permit reconstruction of the raw RTs.

Word reading. Table 1 presents the word reading data—mean RTs in milliseconds and mean proportions of errors—as a function of the within-subjects variable day (1 vs. 5), and the between-subjects variable task (integrated vs. separated). Separate mixed 2×2 ANOVAs were conducted on the control condition means and on the facilitation data and the interference data. Owing to the virtually total absence of errors in word reading, no corresponding analyses are reported for the error data.

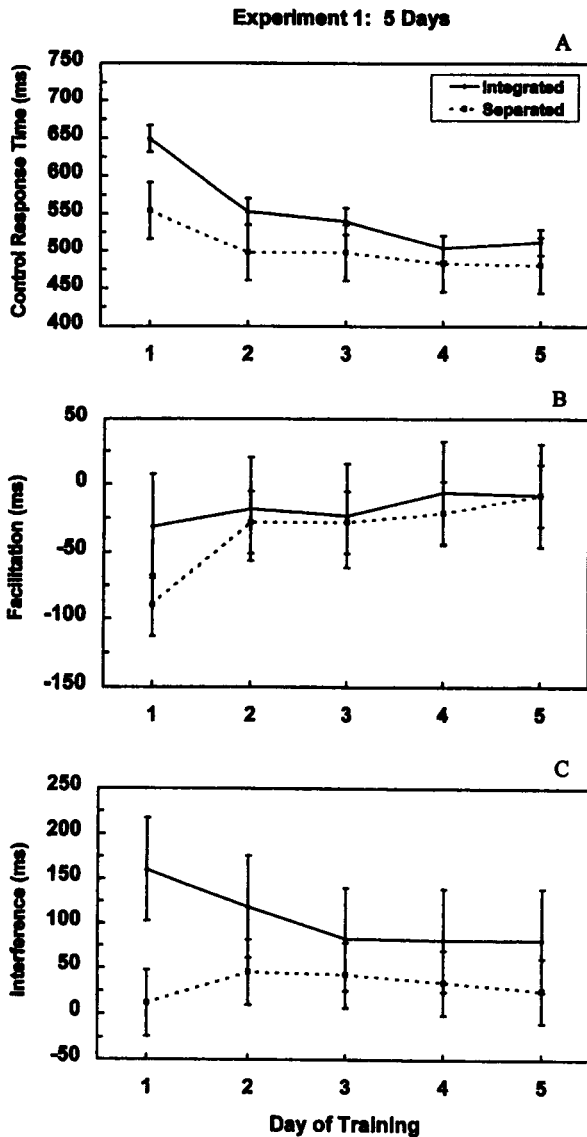


Figure 1. The color naming data for Experiment 1 as a function of day of training and task. Error bars are 95% confidence limits for each task, calculated as set out by Loftus and Masson (1994, Equation 2, p. 482). Panel A shows mean response time (in milliseconds) in the control (XXXX) condition; panel B shows mean facilitation (in milliseconds); panel C shows mean interference (in milliseconds). The facilitation (panel B) and interference (panel C) difference scores were derived by subtraction of the appropriate day control mean (panel A) in such a way that facilitation should be negative and interference positive.

For the control means, there was a marginally significant effect of task [$F(1,6) = 5.36, MS_e = 4,544.12, .05 < p < .10$], indicating that responses in the separated task were overall slightly faster than those in the integrated task, a consistent result across almost all measures in both experiments. Neither the effect of day nor its interaction with task was significant (both $F_s < 1$). Clearly, training on the color naming task had little overall impact on performance in the word reading task.

For facilitation, neither the main effect of task [$F(1,6) = 2.32, MS_e = 343.98, p > .10$] nor that of day [$F(1,6) = 4.35, MS_e = 124.40, .05 < p < .10$] was significant, nor was their interaction [$F(1,6) = 2.98, p > .10$]. Essentially, all values were close to zero, but there was a slight tendency for the small amount of facilitation on Day 1 ($M = 10$ msec) to disappear by Day 5 ($M = -1$ msec).

For interference, there was again a nonsignificant effect of task ($F < 1$) and a nonsignificant interaction of day \times task [$F(1,6) = 2.61, MS_e = 239.83, p > .10$], but here there was a significant effect of day [$F(1,6) = 8.82, p < .05$]. Interference in word reading did rise slightly between Day 1 ($M = -9$ msec) and Day 5 ($M = 14$ msec). This small change was no doubt partly due to the inexplicable negative starting point, so I would not assign much confidence to this finding.

As expected, word reading showed no effect of color compatibility before training. However, studies by Stroop (1935, Experiment 3) and others (e.g., Dulaney & Rogers, 1994) suggested that word reading performance could be altered by training on color naming. On the basis of these studies, it was anticipated that both interference and facilitation might emerge with training. This did not happen. The fact that training in the color naming task hardly influences word reading is a testament to the extreme degree of automaticity attached to word reading.

Color naming. The error data and corresponding error analyses for the 5 days of color naming training are included in Appendix A. Error rates were virtually always under 5% and no systematic pattern was observed, except for the anticipated consistently higher error rates in the incongruent condition. The error data also are reassuring in that there was no speed-accuracy tradeoff.

Panel A of Figure 1 displays the mean RTs for the neutral control condition over the 5 days of training in each of the two tasks. A 2×5 (task by day) mixed ANOVA revealed significant main effects of task [$F(1,6) = 11.90, MS_e = 22,570.30, p < .05$] and of day [$F(4,24) = 12.62, MS_e = 729.28, p < .001$], but a nonsignificant interaction [$F(4,24) = 1.72, p > .10$]. The RT improvements over days of 76 msec for the separated task and 106 msec for the integrated task can be seen as general taskwide improvements. These control data permit reconstruction of the RTs for the congruent and incongruent conditions, but also show that practice affects the control condition, too.

Panel B of Figure 1 displays facilitation. Here, only the main effect of day was significant [$F(4,24) = 3.92, MS_e = 851.87, p < .05$], with facilitation decreasing over days. Neither the main effect of task nor the interaction of day \times task was significant (both $F_s < 1.47$). Despite the absence of a reliable interaction, it is clear that the divergent facilitation on Day 1 contributed disproportionately to the significant reduction over days. If Day 1 is dropped, the effect of day is no longer significant [$F(3,18) = 1.28, MS_e = 414.16, p > .10$]. Why Day 1 should be unique is not obvious, but a similar pattern will be evident in Experiment 2. In sum, facilitation in Experiment 1 was very small and stable in both tasks after the 1st day, averaging about 14 msec for the integrated task and 21 msec for the separated task. Training had little impact on facilitation.

The picture was quite different for interference, shown in panel C of Figure 1. There, only the main effect of task was significant [$F(1,6) = 9.18$, $MS_e = 5,822.23$, $p < .05$]. The effect of day did not approach significance ($F < 1$). The interaction, though, did approach significance [$F(4,24) = 1.98$, $MS_e = 1,924.90$, $p = .13$] and is suggestive given the small sample size. It appears that the tasks differed only on the first 2 days. For the integrated task, interference fell precipitously at first from 160 msec on Day 1 to 83 msec on Day 3, after which it stabilized. For the separated task, interference was unexpectedly low on Day 1 but then rose on Day 2 and declined gradually from 43 msec on Day 3 to 26 msec on Day 5. Mean interference over the last 3 days was about twice as great for the integrated task (83 msec) as for the separated task (34 msec). These differences in interference between the two tasks will be explored more fully in Experiment 2, where they also appeared.

In summary, for a small-sample study with relatively little training, the results were quite clear and also hinted at intriguing task differences. First, contrary to expectation based on prior studies (e.g., Dulaney & Rogers, 1994; Stroop, 1935, Experiment 3) there was very little effect of color naming training on word reading. This finding will be confirmed in Experiment 2. A possible reason for the discrepancy with previous findings will be considered in the General Discussion.

In color naming itself, facilitation was small but reliable, affected little by training or task. Interference, in contrast, was greater than facilitation for both tasks, but especially for the integrated task. Interference fell sharply for the first 2 or 3 days of the integrated task, and then leveled off. From Day 3 on, the integrated task evidenced about twice as much interference as the separated task, with both showing a slow, gradual decline. Experiment 2 will pursue the suggestion that something important and unique is learned on the first few days of training with the integrated version of the task, and that the conflict in the integrated task is also generally harder for subjects to handle.

EXPERIMENT 2

Three principal changes were made in Experiment 2 to increase the amount of training and the statistical power. For power, sample size was increased from 4 to 6 subjects in each task group. For training, subjects completed 288 trials per session, almost tripling the daily training in Experiment 1 (in which there were 108 trials per session). Furthermore, training was extended from 5 to 10 days to expand investigation of the time courses of interference and facilitation.

Method

Subjects. Twelve students from the same pool were paid \$60 each following debriefing after the study. Six subjects were assigned to the integrated task and 6 to the separated task.

Apparatus and experimental control. Testing was carried out on IBM-compatible 286 computers equipped with Tatung CM-1496 14-in. VGA color monitors. When a subject spoke into a Realistic Highball-7 microphone connected to a Realistic SA-150 amplifier,

the signal was sent to a modified keyboard, where it was received as if the hyphen key had been pressed. Programming was done in QuickBASIC 4.5 with millisecond accuracy timing routines taken from Graves and Bradley (1988). Screen background color was black; instructions and control stimuli were presented in white.

Tasks. For the 6 subjects in the integrated task, the word appeared in standard lowercase font in a particular color (red, blue, green, or yellow) at the center of the screen on line 12. For the 6 who did the separated task, a row of asterisks appeared in color on line 12 centered directly above the word, which was again in standard lowercase font and appeared in neutral white on line 13.

Procedure. Instructions were as in Experiment 1. The first task on Day 1 was word reading. All stimuli were presented in lowercase letters at the center of the screen. The words were RED, BLUE, GREEN, and YELLOW. The procedure for the 36 practice trials and the 288 experimental trials followed the same format as Experiment 1, but there were now 96 trials each in the congruent, incongruent, and control conditions during the experimental session. Subjects were allowed to rest briefly halfway.

Each trial began with a 750-msec blank screen. Then the color-word stimulus appeared at the center of the screen and stayed on until the subject responded orally. A 250-msec blank screen followed stimulus offset. Then a message indicated the correct response, and subjects scored their own data as in Experiment 1. The experimenter was present during practice in case subjects had any problems.

The color naming training trials immediately followed the word reading trials on Day 1. Subjects again did 36 practice trials before the 288 experimental trials. The control condition was now the stimulus XXXXX printed in an ink color. On Days 2-9, subjects simply did 288 color naming trials each day. Prior to most sessions, the experimenter reminded the subject to try to respond as quickly as possible without errors. To encourage improvement, at the end of each session subjects saw the message: "Your average time today was n msec per trial. This is the time you should try to beat tomorrow." Accuracy feedback was shown in the following form: "Today you made n mistakes, about $p\%$." This was followed by one of two statements: for 5 to 19 errors, "This error rate is acceptable"; for fewer than 5 errors, "This error rate is very low. Good work."

On Day 10, subjects did a final session of 288 color naming trials and then returned to word reading trials. As on Day 1, 36 practice trials preceded the 288 experimental word reading trials.

Results and Discussion

Data preparation was carried out exactly as in Experiment 1.⁴

Word reading. Table 2 presents the word reading control condition and difference score data—mean RTs and mean proportions of errors—as a function of day (1 vs. 10) and task (integrated vs. separated). Separate mixed 2×2 ANOVAs were conducted on the control condition means and on the facilitation data and the interference data. Corresponding analyses conducted on the virtually nonexistent error rates produced no reliable results and will not be presented.

For the control means, there was no effect of task [$F(1,10) = 1.08$, $MS_e = 1,559.58$, $p > .10$]: Word reading RTs were equivalent in the separated and integrated tasks, as Table 2 shows. However, the reliable main effect of day [$F(1,10) = 66.48$, $MS_e = 372.58$, $p < .001$], coupled with the trend toward an interaction [$F(1,10) = 3.27$, $p = .10$], indicated a benefit due to training for both tasks, but more so for the separated task. Whereas 5 days of training at color naming in Experiment 1 did not affect word reading in the con-

Table 2
Experiment 2, Word Reading: Mean Facilitation, Interference, and Control Response Times (RT, in Milliseconds) and Corresponding Mean Proportions of Errors (PE) as a Function of Day of Training for Both the Integrated and the Separated Tasks

Condition	Task			
	Integrated		Separated	
	RT	PE	RT	PE
	Day 1			
Facilitation	3	-.016	-12	-.002
Interference	0	-.005	-9	.004
Control	455	.016	452	.005
	Day 10			
Facilitation	-4	-.004	22	.002
Interference	7	.007	11	.005
Control	405	.005	374	.002

Note—The control condition values are actual means. The facilitation and interference values are deviations from the corresponding control condition, in which a negative value should accompany facilitation (i.e., faster than control) and a positive value should accompany interference (i.e., slower than control).

control condition, 10 days of color naming training in Experiment 2 did.

For facilitation, neither of the main effects was significant (both $F_s < 1.91$). The marginally significant interaction of day \times task [$F(1,10) = 4.55, MS_e = 568.01, .05 < p < .10$] reflects the anomalous *interference* displayed in the congruent condition for subjects on Day 10 of the separated task. Plainly stated, there was no facilitation in word reading before or after training in color naming.

For interference, there were nonsignificant effects of task and of day \times task (both $F_s < 1$), but only a marginally significant effect of day [$F(1,10) = 4.14, MS_e = 257.62, .05 < p < .10$]. Interference in word reading rose slightly from Day 1 (-4 msec) to Day 10 (9 msec), but this small change was no doubt partly due to the negative starting point. This pattern matches that in Experiment 1.

Previous studies (e.g., Dulaney & Rogers, 1994; Stroop, 1935, Experiment 3) led initially to the expectation that word reading would be affected by color naming training. Clearly, Experiment 2 replicated Experiment 1 in showing no such transition despite overall color naming practice increasing by about a factor of 6 (twice the number of days and three times the number of trials per day). The extreme automaticity of word reading is very difficult to overcome. The fleeting effects reported by previous investigators (e.g., Stroop, 1935, Experiment 3) may be isolated to the multiple-item version of the task, perhaps having to do with screening out surrounding competing stimuli.

Color naming. The color naming error data are displayed in Appendix B with the corresponding ANOVAs. As in Experiment 1, the very low error rates resulted in no apparent effect of days of training or of task. Again, the only observable effect in errors was that the incongruent condition consistently showed the highest error rate, in keeping with the interference in the RT data, to which I now turn.

Panel A of Figure 2 displays the mean RTs for the neutral control condition. A 2×10 (task \times day) mixed ANOVA

revealed significant main effects of task [$F(1,10) = 6.85, MS_e = 9,166.88, p < .05$] and of day [$F(9,90) = 8.97, MS_e = 2,551.72, p < .001$], but a nonsignificant interaction [$F(9,90) = 1.21, p > .10$]. The improvements in RT of about 110 msec for the separated task and about 171 msec for the integrated task can be seen as general taskwide improvements with practice.

Panel B of Figure 2 displays facilitation. The effect of day was not significant ($F < 1$). However, there was a reliable effect of task [$F(1,10) = 6.49, MS_e = 331.60, p < .05$] and a reliable interaction of day \times task [$F(9,90) = 3.09, MS_e = 297.16, p < .01$]. The figure shows that these two effects resulted from a large divergence (for both tasks) on Day 1 relative to subsequent days. In fact, removing the Day 1 data eliminates both effects ($F_s < 1.57$). For unknown reasons, facilitation was highly variable on Day 1 here, a finding also seen in Experiment 1. Once subjects settled down after Day 1, both tasks showed consistently small and equivalent amounts of facilitation. Facilitation was hardly affected by training, again in keeping with Experiment 1.

The pattern was different for interference, displayed in panel C of Figure 2. Both main effects were significant: task [$F(1,10) = 7.70, MS_e = 6,973.97, p < .05$] and day [$F(9,90) = 10.21, MS_e = 492.84, p < .001$], as was the interaction [$F(9,90) = 5.76, p < .001$]. On this basis, the tasks were analyzed separately. For the separated task, although the effect of day was not significant [$F(9,45) = 1.34, MS_e = 496.42, p > .10$], there was a marginally significant linear trend [$F(1,5) = 4.68, MS_e = 1,041.16, .05 < p < .10$]. Interference declined gradually and regularly from 46 to 25 msec over the 10 days. On the other hand, for the integrated task, the effect of day was highly significant [$F(9,45) = 14.73, MS_e = 489.26, p < .001$]. There was a significant linear trend [$F(1,5) = 17.62, MS_e = 1,988.56, p < .01$] and a significant quadratic trend [$F(1,5) = 91.14, MS_e = 117.24, p < .001$]. Interference fell quite precipitously at first, from 167 msec on Day 1 to 86 msec on Day 4, and then gradually declined to 56 msec by Day 10.⁵ Note the close correspondence to Experiment 1.

What is different about the two tasks is interference over the first 2–3 days, as suggested by Experiment 1. Indeed, if the first 3 days are removed and the two tasks are compared only for Days 4–10, panel C of Figure 2 suggests a very similar pattern. This is supported by a 2×7 ANOVA, which showed reliable main effects of task [$F(1,10) = 6.18, MS_e = 4,072.15, p < .05$] and of day [$F(6,60) = 3.55, MS_e = 277.62, p < .01$], but a nonsignificant interaction ($F < 1$).

In sum, the results were quite thoroughly consistent with those of Experiment 1, but were more steady given the increased training and number of subjects performing each task. Again, word reading was unaffected by training on color naming. In color naming, facilitation was once again small and reliable, showing little change with training. Interference, in contrast, was considerably larger than facilitation, and the two tasks showed different training patterns. This reinforces the earlier claim that something

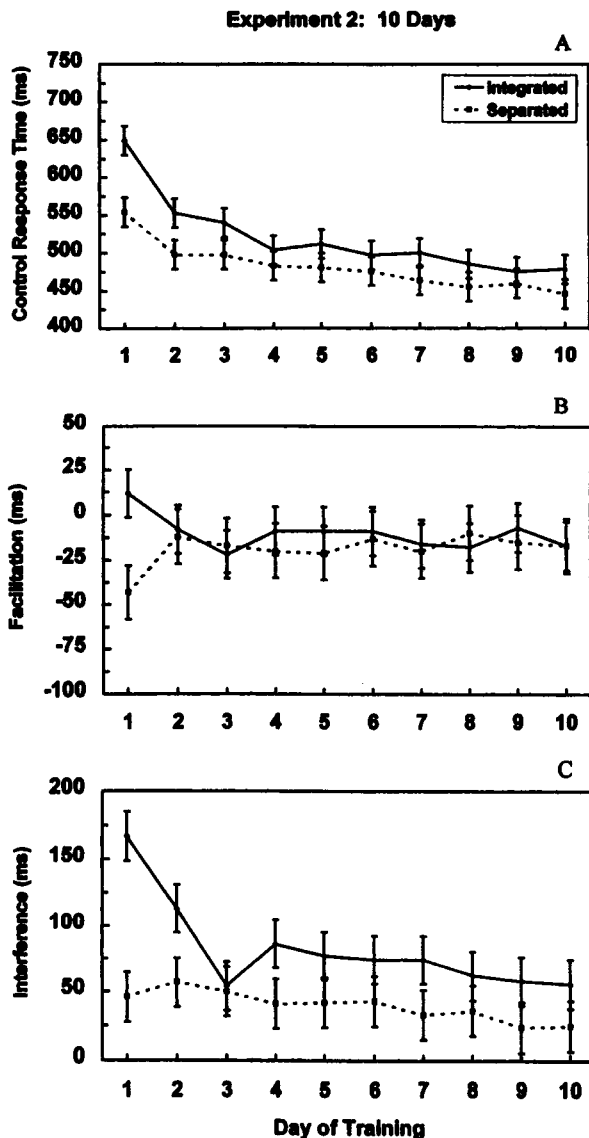


Figure 2. The color naming data for Experiment 2 as a function of day of training and task. Error bars are 95% confidence limits for each task, calculated as set out by Loftus and Masson (1994, Equation 2, p. 482). Panel A shows the response times (in milliseconds) in the control (xxxx) condition; panel B shows mean facilitation (in milliseconds); panel C shows mean interference (in milliseconds). The facilitation (panel B) and interference (panel C) difference scores were derived by subtraction of the appropriate day control mean (panel A) in such a way that facilitation should be negative and interference positive.

special must have to be learned early in training on the integrated version of the task.

GENERAL DISCUSSION

Both experiments support the same conclusions. First, training in the Stroop color naming task led to a marked reduction in interference for both the integrated and the

separated versions of the task. Over 5 days in Experiment 1, interference fell 49% (from 160 to 82 msec) for the integrated stimuli and 43% (from 46 to 26 msec) for the separated stimuli.⁶ Over 10 days in Experiment 2, interference fell 66% (from 167 to 56 msec) for the integrated stimuli and 46% (from 46 to 25 msec) for the separated stimuli. The magnitude of the interference decline is in keeping with previous work (Dulaney & Rogers, 1994; Stroop, 1935). Note, though, that there was still reliable interference even after extensive training in which subjects were encouraged to overcome interference in any way they could.

Second, by including a congruent condition, this study permitted the comparisons done on interference to be done on facilitation as well. Disregarding the noisiness of the 1st day evident in both experiments, facilitation was essentially constant at 15–20 msec, entirely unaffected by training. This was true for both integrated and separated stimuli over both 5-day and 10-day training sessions.

Third, there was no evidence of any impact whatsoever due to color naming training on performance in the word reading task. The absence of interference conflicts with prior work (Dulaney & Rogers, 1994; Stroop, 1935, Experiment 3), in which color naming practice led to a “reverse Stroop” effect—albeit a quite temporary one—in word reading, perhaps reflecting a difference between the multiple-item and single-item procedures. These previous studies did not measure facilitation.

Training and Separation Versus Integration

The first question raised in the introduction concerned the degree to which the interference pattern over training in color naming in the single-item procedure used here corresponds with that in the older, multiple-item procedure. Restricting attention to integrated stimuli, there appears to be a quite good correspondence between the present results and those of Stroop (1935, Experiment 3) and Dulaney and Rogers (1994). Interference is very strong initially, falls quickly over the first couple of days, and then settles into a more gradual decline. The only difference between the multiple-item and single-item procedures is in the overall magnitude of interference, which appears larger in the multiple-item version. The patterns of training effects, however, are very similar for the two versions.

The pattern for separated stimuli is quite different. Here, interference was much smaller at the outset and declined gradually, without the sharp early drop. In fact, after the first 3 days, although interference was about twice as large in the integrated case as in the separated case, the two tasks showed similar rates of decline. These data suggest that there are two contributions to interference for integrated stimuli: a more dramatic but more fleeting cost due to integration, and a smaller but more resistant cost stemming from the existence of two conflicting responses. Subjects must first overcome the integration problem, which they do quite quickly, and can then begin to “chip away” at the problem of responses deriving from two dimensions. If correct, this admittedly post hoc

hypothesis has a clear theoretical implication: Stroop interference, at least in the traditional integrated task, may not be caused by a single process. There is also a "practical" implication: Researchers employing this version of the task should consider training their subjects, given the dramatic changes in the early going.

It also appears that the effect of training in color naming spills over into word reading for the multiple-item version but not for the single-item version. Actually, there are three differences between the two versions. The most obvious difference—different numbers of stimuli (one or many) are in view at the same time—may not be the critical one. In the single-item version, control and incongruent trials are mixed; they are separated in the multiple-item version. Moreover, in the present single-item version, congruent trials are also mixed in. Mixing trial types may disrupt whatever process is responsible for the "reverse Stroop" interference observed after color naming training in multiple-item experiments.

Facilitation Versus Interference

Perhaps the most compelling result of the present study is that neither integration/separation nor amount of training influenced facilitation, but both profoundly affected interference. The two derived measures appear to be decoupled, a finding with considerable precedent in the literature, as outlined in the introduction. Moreover, MacLeod and MacDonald (1997) directly supported an inadvertent reading hypothesis for facilitation, an account independent of the prevalent explanations of interference. In explaining interference, all existing theories allow for facilitation to arise via the same mechanism(s). Cooperation between the word and color produces facilitation in the same way that competition produces interference. As documented by MacLeod and MacDonald, this is true for the major models in the literature—automaticity accounts (Logan, 1980), connectionist accounts (Cohen et al., 1990; Phaf et al., 1990), network accounts (W. R. Glaser & M. Glaser, 1989), and process-dissociation-based accounts (Lindsay & Jacoby, 1994). Yet there is mounting evidence against this sort of joint explanation, and the present experiments add to that body of evidence. If facilitation does not arise from the same causes as interference, then any theory claiming that it does so must be wrong. Intriguingly, all of the existing theories do make such a claim.

In sum, the two experiments described here have documented how two fundamental variables—integration and training—affect performance in the Stroop task. These findings pose new theoretical challenges for investigators interested in this longstanding attentional puzzle. It appears that we may need distinct mechanisms to explain facilitation and interference, given that they respond differently to a variety of manipulations. Indeed, explaining interference alone, at least in the traditional integrated color-word task, may not be possible by using a single process, given the sharp transition in interference between early and later days of training. Existing theories do not readily accommodate such suggestions. Even 60 years on, then, we remain rather distant from a full understanding of

Stroop's (1935) deceptively simple appearing phenomenon, and the phenomenon in turn remains an important testing ground for our theories of attention.

REFERENCES

- ACKERMAN, P. L., & SCHNEIDER, W. (1984, November). *Practice effects and a model for Stroop interference*. Paper presented at the annual meeting of the Psychonomic Society, San Antonio.
- BROWN, A. S., ENGLE, C. A., & JONES, T. C. (1992, November). *Incrementing interference in the Stroop color-word task*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis.
- CATTELL, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63-65.
- CHENG, P. W., & PACHELLA, R. G. (1984). A psychophysical approach to dimensional separability. *Cognitive Psychology*, *16*, 279-304.
- COHEN, J. D., DUNBAR, K., & MCCLELLAND, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332-361.
- DULANEY, C. L., & ROGERS, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 470-484.
- DUNBAR, K., & MACLEOD, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 622-639.
- DUNN, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, *9*, 242-257.
- DYER, F. N. (1971). The duration of word meaning responses: Stroop interference for different preexposures of the word. *Psychonomic Science*, *25*, 229-231.
- DYER, F. N. (1973). Interference and facilitation for color naming with separate bilateral presentations of the word and color. *Journal of Experimental Psychology*, *99*, 314-317.
- DYER, F. N., & SEVERANCE, L. J. (1973). Stroop interference with successive presentations of separate incongruent words and colors. *Journal of Experimental Psychology*, *98*, 438-439.
- ELLIS, N. R., & DULANEY, C. L. (1991). Further evidence for cognitive inertia in persons with mental retardation. *American Journal on Mental Retardation*, *95*, 613-621.
- FLOWERS, J. H., & STOUPE, C. M. (1977). Selective attention between words, shapes and colors in speeded classification and vocalization tasks. *Memory & Cognition*, *5*, 299-307.
- GARDNER, R. W., & LONG, R. I. (1962). Cognitive controls of attention and inhibition: A study of individual consistencies. *British Journal of Psychology*, *53*, 381-388.
- GARNER, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- GARNER, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.
- GATTI, S. V., & EGETH, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society*, *11*, 181-184.
- GLASER, M. O., & GLASER, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception & Performance*, *8*, 875-894.
- GLASER, W. R., & GLASER, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*, 13-42.
- GRAVES, R., & BRADLEY, R. (1988). More on millisecond timing and tachistoscope applications for the IBM PC. *Behavior Research Methods, Instruments, & Computers*, *20*, 408-412.
- KAHNEMAN, D., & CHAJCZYK, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, *9*, 497-509.
- KAHNEMAN, D., & HENIK, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.
- LINDSAY, D. S., & JACOBY, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 219-234.

- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.
- LOGAN, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, *12*, 523-553.
- LOGAN, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, *39*, 367-386.
- LOWE, D., & MITTERER, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology*, *36*, 684-700.
- LUPKER, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*, 485-495.
- MACLEOD, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203.
- MACLEOD, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, *121*, 12-14.
- MACLEOD, C. M., & DUNBAR, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 126-135.
- MACLEOD, C. M., & HODDER, S. L. (1998). Presenting two incongruent color words on a single trial does not alter Stroop interference. *Memory & Cognition*, *26*, 212-219.
- MACLEOD, C. M., & MACDONALD, P. A. (1997). *Facilitation in the Stroop task is illusory: The inadvertent reading hypothesis*. Manuscript submitted for publication.
- MONAHAN, J. S., & LOCKHEAD, G. R. (1977). Identification of integral stimuli. *Journal of Experimental Psychology: General*, *106*, 94-110.
- PHAF, R. H., VAN DER HEIJDEN, A. H. C., & HUDSON, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, *22*, 273-341.
- SHIFFRIN, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (pp. 739-811). New York: Wiley.
- SHOR, R. E., HATCH, R. P., HUDSON, L. J., LANDRIGAN, D. T., & SHAFER, H. J. (1972). Effect of practice on a Stroop-like spatial directions task. *Journal of Experimental Psychology*, *94*, 168-172.
- SIMON, J. R., CRAFT, J. L., & WEBSTER, J. B. (1973). Reactions toward the stimulus source: Analysis of correct responses and errors over a 5-day period. *Journal of Experimental Psychology*, *101*, 175-178.
- STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.
- TZELGOV, J., HENIK, A., & BERGER, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*, 727-735.
- VANAYAN, M. (1992). *Relating interference and facilitation in the Stroop task: An individual-differences approach*. Unpublished doctoral dissertation, University of Toronto, Scarborough.
- YEE, P. L., & HUNT, E. (1991). Individual differences in Stroop dilution: Tests of the attention-capture hypothesis. *Journal of Experimental Psychology: Human Perception & Performance*, *17*, 715-725.

NOTES

1. Following Dulaney and Rogers (1994), I have transformed Stroop's (1935) "whole card" times to "per item" times by simply dividing the total time per card by the number of items per card. This procedure probably makes the decline in interference appear more precipitous because Stroop had subjects correct errors, thereby inflating interference, and because we do not know if errors declined over days.
2. The magnitude of interference obtained by Stroop (1935, Experiment 3) and by Dulaney and Rogers (1994) is extraordinarily large. This may reflect factors operating in the multiple-item card procedure but not in the single-item computer version (cf. Brown, Engle, & Jones, 1992), or it may reflect the possibly questionable derivation of a "per word" time from a single total time over multiple items.
3. In preparation for a training study involving sessions of similar length (MacLeod & Dunbar, 1988), we compared subject-scored to experimenter-scored accuracy for a small group of subjects. The number and distribution of errors across conditions were very similar regardless of who scored accuracy, so we had the subject score his/her own errors to eliminate an extremely tedious task for the experimenter.
4. For 3 subjects, data from a single training session (one on Day 7, two on Day 9) were lost due to disk writing errors. To analyze the training data, missing values were estimated as the average of the preceding and the succeeding day values for the corresponding conditions for that subject. As will become apparent, the orderliness of the data suggests that this interpolation was not inappropriate.
5. In Experiment 2, 3 of the 6 subjects in the integrated task showed surprisingly little interference on Day 3, but returned to more predictable levels on Day 4. Comparison with Experiment 1 suggests that the Day 3 point here is anomalous, so I have shifted to using Day 4 as the point of comparison.
6. In fact, Day 1 interference for separated stimuli in Experiment 1 was an outlier at 12 msec. Because the Day 2 value of 46 msec was the same as in Experiment 2, with a similar decline over subsequent days, the Day 2 value is used here. There is certainly no reason to believe that interference is nonmonotonic.

APPENDIX A

Experiment 1: Proportions of Errors in Color Naming Over the 5 Days of Training in the Congruent (CG), Incongruent (IN), and Control (CT) Conditions as a Function of Task (Integrated vs. Separated)

Condition	Day														
	1			2			3			4			5		
	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT
Integrated	.01	.09	.00	.00	.03	.01	.01	.04	.00	.00	.02	.00	.00	.02	.00
Separated	.00	.02	.01	.00	.03	.00	.00	.04	.00	.00	.01	.01	.00	.02	.00

Note—Separate 2×5 (task \times day) ANOVAs of the two principal measures—facilitation (congruent minus control) and interference (incongruent minus control)—revealed no significant main effects or interactions (all six F s < 1.74). Basically, the error rates were very low and unaffected by the manipulations. It is noteworthy, though, that the incongruent condition in both tasks produced elevated error rates on every day, consistent with the interference observed in the response time data.

APPENDIX B
Experiment 2: Proportions of Errors in Color Naming Over the
10 Days of Training in the Congruent (CG), Incongruent (IN), and
Control (CT) Conditions as a Function of Task (Integrated vs. Separated)

Condition	Day																																
	1			2			3			4			5			6			7			8			9			10					
	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT	CG	IN	CT
Integrated	.00	.09	.00	.00	.07	.01	.00	.06	.02	.00	.04	.00	.00	.06	.01	.00	.07	.01	.00	.05	.00	.00	.07	.01	.00	.06	.01	.00	.08	.01	.00	.08	.01
Separated	.01	.06	.01	.01	.05	.01	.01	.04	.00	.01	.06	.01	.01	.06	.01	.00	.05	.01	.02	.06	.01	.02	.09	.01	.01	.07	.02	.01	.07	.01	.01	.07	.01

Note—Separate 2×10 (task \times day) ANOVAs of facilitation and interference revealed no significant effects (all $F_s < 1.35$) except for the main effect of task for the facilitation data [$F(1,10) = 7.02, MS_e = .0003, p < .05$], reflecting slightly greater facilitation overall in the integrated task, a result not evident in Experiment 1. Otherwise, the facilitation error rates were very low and did not differ. Again, the only consistent pattern was that the incongruent condition showed substantially higher error rates than the control and congruent conditions, corresponding to the interference observed in the response time data.

(Manuscript received September 22, 1997;
 revision accepted for publication October 6, 1997.)