Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 1, 126–135

0278-7393/88/$00.75

Training and Stroop-Like Interference: Evidence for a Continuum of Automaticity

Colin M. MacLeod
University of Toronto, Scarborough, Ontario, Canada

Kevin Dunbar
Carnegie-Mellon University

Three experiments varied the extent of practice in an analog of the Stroop color-word task. Each experiment involved four phases: (a) baseline naming of four familiar colors, (b) training in consistently naming four novel shapes by using the names of the same four colors, (c) naming the colors when they appeared in the form of the shapes, and (d) naming the shapes when they appeared in color. In Experiment 1, with up to 2 hr of training in shape naming, colors were named much faster than shapes. Interference was observed only in Phase 4. In Experiment 2, with 5 hr of training, shape naming sped up, but was still slower than color naming. Nevertheless, there was symmetrical interference in Phases 3 and 4, and this persisted 3 months later without further training. Experiment 3 replicated this pattern and then extended practice to 20 hr, by which time shape and color naming were equally rapid. After 20 hr, interference appeared only in Phase 3, reversing the original asymmetry. The overall pattern is inconsistent with a simple speed of processing account of interference. The alternative idea of a continuum of automaticity—a direct consequence of training—remains plausible, and the implications of this perspective are considered.

The Stroop task is one of the most familiar experimental demonstrations in the domain of cognitive psychology. The basic situation, first described by Stroop (1935b), is this: A color word such as GREEN appears in an ink color such as red. If the subject's task is to read the word and ignore the color (i.e., say "green"), there is no evidence of difficulty relative to reading the word in standard black ink. However, if the subject's task is to name the ink color and ignore the word (i.e, say "red"), there is considerable difficulty relative to naming a color patch. This is the phenomenon of *Stroop interference*, and the particularly intriguing aspect of the phenomenon is its *asymmetry*. The word interferes with naming the color, but the color does not interfere with reading the word.

A major reason why this task is a favorite demonstration is that there are two simple and compelling explanations for the phenomenon. Consider the *relative speed of processing* hypothesis first. This account has been offered quite frequently (e.g., Dyer, 1973; Morton & Chambers, 1973). The argument is that the two processes—word reading and color naming—are carried out in parallel, and that word reading is accomplished faster. It is assumed that the faster process can interfere with the slower one, but not vice versa. Thus, the relative speed of processing account (often called the "horse-race" model), handles the basic Stroop asymmetry. In fact, as Dyer (1973) maintained in his review of the literature, this account also appears to handle much of the other available data.

The second explanation, the *automaticity* account, has been with us since the beginning of work in the area (see, e.g., Cattell, 1886, p. 65; James, 1890, p. 559). When this view began reappearing more recently (e.g., Posner & Snyder, 1975), the idea was that certain processes were automatic by virtue of being rapid, independent of processing strategy, and not reliant upon cognitive resources. Furthermore, automaticity was "all or none"; either a process was automatic or it was not automatic (e.g., Hasher & Zacks, 1979; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Reading an isolated word would be a prototypical example of an automatic process. Other processes were "controlled," in that they were slower, depended on processing strategy, and required cognitive resources. Color naming might well fall into this category. The basic idea was that automatic processes could not be suspended and would therefore produce interference with controlled processes, but not vice versa. Thus, the automaticity account explains the basic Stroop asymmetry, as it appears to handle other findings in the literature.

Although widely accepted in the literature, each of these two explanations has come under considerable criticism in the past few years. We will present only two illustrations here, but there are others (e.g., Glaser & Glaser, 1982). Dunbar and MacLeod (1984) called into question the straightforward relative speed of processing hypothesis. In our studies, we used the standard version of the Stroop task but rotated the color words (e.g., upside down and backward). As we expected, reading of such words was dramatically slower than reading normal words. Yet, even when word reading was far slower than color naming, incongruent words still interfered with

This research was supported by Grants A7459 and E6532 from the Natural Sciences and Engineering Research Council of Canada to Colin M. MacLeod.

We wish to thank P. J. Lootsteen for carrying out Experiments 1 and 2, and Steve Turner and P. J. Lootsteen for programming assistance. We are also grateful to the subjects of Experiment 3 for their patience and good humor in the face of a tedious task. In addition, we thank H. Roediger, J. Neely and two anonymous reviewers for their helpful comments on a previous version.

Kevin Dunbar is now at McGill University.

Correspondence concerning this article should be addressed to Colin M. MacLeod, Division of Life Sciences, University of Toronto, Scarborough Campus, Scarborough, Ontario, Canada M1C 1A4.

color naming, and to the same extent as in the normal orientation. Put another way, the slower process was affecting the faster one, a finding inconsistent with the relative speed of processing account.

Kahneman and Chajczyk (1983) called into question the automaticity view with their demonstration of "dilution" of interference. Using a modification of the Stroop task in which the color word appeared below a color patch, they observed that an additional word in the display actually reduced the interference when naming the color patch. Apparently, another dimension of the stimulus display could split attention, thereby lessening the impact of the color word on naming of the color patch. This dilution is inconsistent with the idea of automatic reading; attentional allocation should not affect the amount of interference observed if the process is automatic.

Kahneman and Chajczyk (1983) took their data as evidence against a strong version of automaticity and opted for what they called a weak version. In the strong version, processes are either automatic or they are not automatic. Such a dichotomy seems at odds with their evidence, so a gradient of automaticity may be more appropriate. This view, which we will characterize as the *continuum of automaticity* view, is shared by other contemporary theorists. Logan (1985, p. 371) says "there is clear evidence that practice is important in producing automaticity, which suggests that automaticity may be learned . . . . The idea that automaticity is learned and that attributions of automaticity are relative judgments suggests that automaticity should be viewed as a continuum." Shiffrin (in press, p. 33) seems to agree: "It is possible that a process is either automatic or not, and that practice merely increases the probability of the former . . . but it seems more likely that performance and automatism both improve gradually with consistent practice."

One of the most obvious testing grounds for theories of the Stroop task is manipulations of practice, or training. This was evident to Stroop (1935b), who examined the issue in his infrequently cited Experiment 3. It is also clear in the comments of Logan and Shiffrin just cited. Furthermore, according to both of the accounts just described, it should be possible to manipulate interference via training. What is the status of the training literature?

In fact, there was a long tradition of training studies in this domain prior to Stroop's (1935b) landmark study. Most were aimed at testing the differential practice hypothesis for the fact that words are read aloud faster than colors can be named aloud. Brown (1915) practiced subjects extensively both on naming colors and on reading color words, predicting that color naming should benefit more from practice than should word reading because color naming is the less practiced skill initially. He found that performance of both tasks improved substantially, yet the ratio of color-naming times to word-reading times remained roughly constant. This result led him to argue that differential practice could not be the cause of the difference between color-naming and word-reading times. In a developmental study (using grades 1 through 9), Ligon (1932) also claimed to observe improvement in both tasks with a constant ratio across age, and reached the same conclusion as Brown. However, contrary to these findings, Lund (1927) presented developmental data that showed a much larger improvement with age in word reading than in color naming, consistent with the differential practice hypothesis. Given Lund's (1927) critique of Brown (1915) and Stroop's (1935a) critique of Ligon (1932), the differential practice hypothesis has remained quite tenable. In fact, Dyer (1973) argued for its continued importance.

Surprisingly, there are few recent efforts to manipulate practice in the context of the Stroop task. All have involved teaching subjects nonsense syllable responses for either colors, words, or both (Glaser & Dolt, 1977; Pritchatt, 1968, Experiment 3; Stirling, 1979, Experiment 2). Although interference has been observed under these conditions, the experiments were not aimed at directly testing the two explanations set out earlier. Yet the two views make predictions that training studies are logically well suited to evaluate. This is the aim of the experiments we will report.

Under the relative speed of processing account, practicing one dimension should speed that dimension relative to the other; if the initially slower dimension is practiced until it becomes the faster one, then the pattern of interference should reverse. Along the way, when the processing times for the two dimensions are roughly equivalent, they should interfere with each other roughly equivalently. Under the continuum of automaticity view, practicing one dimension should lead to increased automatization of that dimension, in turn resulting in increased interference attributable to that dimension when it must be ignored. Note that the continuum of automaticity explanation does not require that interference be directly predictable from processing times. It is quite possible for the dimension that is processed more slowly to interfere with the one that is processed more quickly, as long as the degree of automaticity of the slower dimension is sufficient. This is the main difference in the predictions of the two theories.

We have said nothing so far about facilitation, but it can also be examined in the context of the Stroop task. When a color word is presented in the congruent ink color (e.g., GREEN in green), naming of the ink color may actually be accelerated relative to the color patch control (see Dyer, 1973). Although less attention has been given to the explanation of facilitation in the literature—probably because it is nowhere near as robust as interference—both theories do make predictions. Under the relative speed of processing account, facilitation derives from the faster dimension assisting the slower dimension when they are congruent. Thus, facilitation should occur only where interference does, and it would seem reasonable to expect the amounts of each to be positively correlated. Under the continuum of automaticity account, the prediction is less clear. Still, it would appear that the more automatically a dimension is processed, the greater should be its tendency both to interfere with and to facilitate the processing of the other dimension. Thus, facilitation should not be observed without interference, although it is possible that a process will have to be more highly automated to obtain the less reliable phenomenon of facilitation than it will to obtain the more reliable phenomenon of interference.

Because of the extensive practice that has accrued to word reading in the standard color-word version of the Stroop task, it was necessary to create a version of the task in which practice could be controlled and observed from the outset. To replace words, we selected unfamiliar shapes from a set of random polygons (Vanderplass & Garvin, 1959). We then assigned color names to these shapes and trained subjects in applying these new color names to the shapes. After variable amounts of training, we monitored interference and facilitation by asking subjects to name the shapes when they appeared in color and the colors when they appeared on shapes. The goal was to vary training sufficiently to speed up or to increasingly automatize shape naming (or both), and then to observe whether the patterns of interference and facilitation over practice fit with the predictions of the two theories.

Experiment 1

In our earlier work (Dunbar & MacLeod, 1984), we obtained a full-blown Stroop effect for transformed words when subjects had no practice in processing these novel stimuli. We had not expected the transformed words to interfere with naming the colors from the very beginning of training because the words would be so slow and difficult to read initially. Yet interference did appear right away, so we decided in the present experiments to begin our exploration using small amounts of training. If the effect emerged early in training, we did not want to miss it.

The two dominant theoretical perspectives make different predictions about the outcome of this first experiment. Under the strong version of automaticity, where word reading is automatic but color naming is not in the standard color-word Stroop task (e.g., Posner & Snyder, 1975), we should not observe any interference at all in our modified Stroop task. After all, our task eliminates word reading altogether, using only color naming and shape naming, neither of which is automatic. Facilitation should be absent as well. If there is interference from the outset, this would provide further evidence against the strong automaticity view, but would be quite consistent with the continuum of automaticity explanation. Here, color naming simply is more automatic than shape naming due to their differential learning histories. The goal would then be to manipulate their learning histories over experiments and observe how this affected interference and facilitation.

The relative speed hypothesis, on the other hand, predicts a clear pattern of interference and facilitation. Initially, the familiar ink color names should be processed much faster than the new shape names, so colors should affect shapes but not vice versa. As shape-naming practice progresses, interference and facilitation should start to appear in naming shapes because the processing times of shapes and colors will begin to overlap. Correspondingly, interference and facilitation should diminish in color naming. In the limit, it might even be possible to obtain interference and facilitation only in the color-naming condition when shape naming has been so extensively practiced as to become much faster than color naming.

Our primary concern in Experiment 1 was to discover whether shape name training could eventually lead to the emergence of interference and facilitation in color naming and, if so, at what point in training this would occur. To hedge our bets, we tried several different amounts of shape name training: (a) 16 shape-naming trials only, (b) 192 shape-naming trials, (c) 288 shape-naming trials, and (d) 576 shape-naming trials (288 on each of two successive days). As is apparent, these were chosen to range from very few through to the most that could be done in a 1-hr session on each of 2 days. How would training affect interference when naming colors versus shapes?

Method

Subjects. The 22 subjects were undergraduate volunteers from introductory psychology at the Scarborough Campus of the University of Toronto. There were 6 subjects in each of the 16-trial, 192-trial, and 288-trial groups, and 4 subjects in the 576-trial group. Subjects were assigned randomly to groups and were tested individually.

Apparatus and stimuli. All testing was under the control of an Apple II+ microcomputer with an Electrohome color monitor and Supercolor Board interface. Response times were measured as the interval between stimulus onset and the subject's vocal response into a microphone, which tripped a Lafayette voice key sending an interrupt to the computer. Accuracy was scored by the experimenter who observed every trial and added the accuracy information to the data file.

The shape stimuli were four low-association-value random polygons selected from the set developed by Vanderplass and Garvin (1959), with the restriction that they be highly discriminable from each other. Tiny alterations were introduced in transforming the printed shapes into their graphic counterparts. The colors were green, pink, orange, and blue, also chosen for their ease of discrimination. The four shapes were presented in white when used as control materials in the shape-naming trials and during all shape-training sessions.

Phase 1: Color-naming baseline. Prior to the baseline trials, subjects were shown each of the four colors on a square and told the color's name, in case of ambiguity in identification. After this familiarization, 72 color-naming trials were conducted. Each of the four colors appeared either four or five times on each of the four random shapes (before the shapes were assigned names). Subjects were instructed to ignore the shape and simply to name each color as quickly as possible, avoiding errors.

Phase 2: Shape-name training. At the outset of Phase 2, the four shapes were shown one at a time with their respective color names for study. The newly assigned color names were those of the four colors used in Phase 1, and the same shape-name pairings were used across all subjects and experiments. After studying all pairs one at a time, each pair was tested. A shape was shown and the subject had to provide its name. This study–test cycle was repeated until the subject correctly named all four shapes on two successive trials. Because there were only four shapes, most subjects achieved this criterion in only two trials. All parts of this acquisition were subject paced.

Once subjects knew the shape names, shape-naming practice began. On each trial, one of the shapes was presented in white on a dark background. The subject was to name the shape aloud as rapidly as possible without making mistakes. Each shape was presented an equal number of times in a random sequence. Thus, each shape was named only four times in the 16-trial condition, 48 times in the 192-trial condition, 72 times in the 288-trial condition, and 144 times in the 576-trial condition.

Phase 3: Naming colors on shapes. Phases 3 and 4 used two-dimensional stimuli: The random shapes appeared in color. For half of the subjects in each group, Phase 3 preceded Phase 4; for the other half, the order was reversed. This counterbalancing of order had no

Table 1
Experiment 1: Mean Response Times (in Milliseconds) and Overall Mean Proportions of Errors (E) for the Four Phases of Each Training Condition

No. of training trials	Phase 1: Baseline colors	Phase 2: Baseline shapes	Phase 3: Naming colors on shapes			Phase 4: Naming shapes in color		
			CG	CT	IN	CG	CT	IN
16	606	910	575	572	586	724	754	834
192	657	795	568	573	568	645	687	749
288	700	862	630	619	642	760	818	933
576	577	741	587	612	604	713	768	864
E	.003	.018	.023	.012	.019	.010	.007	.061

Note. The three conditions in Phases 3 and 4 are congruent (CG), control (CT), and incongruent (IN).

apparent effect on the data and will not be discussed further. It is in Phases 3 and 4 that interference and facilitation can be evaluated.

In Phase 3, there were 20 different stimuli: Each of the four colors appeared on each of the four shapes and on a square, the control shape. The 72 trials consisted of three types: (a) 24 control trials where each color appeared on a square (6 times each), (b) 24 congruent trials where each color appeared on only the shape with the corresponding color name (6 times each), and (c) 24 incongruent trials where each color appeared on each shape except the one with the corresponding color name (2 times each). Trial sequence was random, and subjects were instructed to respond as quickly as possible, avoiding errors.

Phase 4: Naming shapes in colors. Again, there were 20 different stimuli: Each of the four shapes could appear in each of the four colors and in white, the control color. The 72 trials consisted of three types: (a) 24 control trials where each shape appeared in white (6 times each), (b) 24 congruent trials where each shape appeared only in the color corresponding to its learned name (6 times each), and (c) 24 incongruent trials where each shape appeared in every color except the one corresponding to its learned name (2 times each). Trial sequence was random and the same speed–accuracy instructions applied.

No feedback was provided at any point in the procedure (except for occasional oral corrections necessary during familiarization at the start of Phase 2). Short relaxation periods were provided between phases. For the 2-day group (576 trials), all phases were conducted on both days, but only data from Day 2 are reported. The data from Day 1 for this group closely replicated those of the 288-trial group.

Results and Discussion

Error rate was very low in all four training groups for all phases of the experiment. Because no group differences in error rate were evident, only the mean proportions of errors averaged over groups are reported in Table 1. The error data showed no evidence of any interference in color naming (Phase 3), but did suggest that interference occurred in shape naming (Phase 4). There was no suggestion of any speed–accuracy tradeoff.

The primary data to be discussed are mean naming times for only the correct responses in each condition. These are displayed separately by phase for each of the four training groups in Table 1. Statistical evaluation of the data basically confirmed the pattern evident in the table.[1] The data of Phases 3 and 4 were analyzed separately because they differ substantially in overall latency.

For the color-naming data (Phase 3), there was no evidence of either interference or facilitation for any amount of training. Whether compatible or incompatible with the color, the shape names had no effect on color naming. A 4 × 3 mixed analysis of variance (ANOVA) with amount of training as the between-subjects variable (16, 192, 288, or 576 trials) and condition as the within-subjects variable (congruent, control, or incongruent) revealed no significant effects, all three Fs < 1.34.

On the other hand, there was consistent evidence of strong interference and modest facilitation in the shape-naming data regardless of amount of training (Phase 4). Incompatible colors slowed shape naming (mean interference of 88 ms), whereas compatible colors speeded shape naming (mean facilitation of 46 ms). A corresponding 4 × 3 mixed ANOVA showed a marginally significant effect of amount of training, $F(3, 18) = 2.81$, $p = .07$, $MS_e = 19{,}775.82$, a highly significant effect of condition, $F(2, 36) = 76.68$, $p < .001$, $MS_e = 1{,}298.63$, and no reliable interaction between the two variables, $F(6, 36) = 1.22$.

In sum, although shape naming consistently showed both facilitation and interference from colors regardless of the extent of training, color naming showed neither interference nor facilitation from shapes even with 2 days of training. Thus, facilitation and interference can occur when neither dimension of the compound stimulus is a word. Yet the strong version of the automaticity hypothesis maintains that the asymmetrical interference in the standard Stroop task is caused by word reading but not color naming being automatic. If the color name is not processed automatically but nevertheless can interfere with and facilitate shape naming, then the strong version of the automaticity hypothesis cannot explain the data.

To salvage the automaticity hypothesis, it must be assumed that automaticity is not all-or-none, but that different tasks fall at different points along a continuum of automaticity (see Kahneman & Chajczyk, 1983; Logan, 1985; Shiffrin, in press). In the standard Stroop color-word task, word reading is more practiced and hence more automatic than color naming, but both processes are nevertheless at least somewhat automatic. It is a small step to extend this logic to the present experiment and assume that color naming here is more automatic than

[1] Although it is reasonable to expect baseline shape-naming times to decline with practice, the variation in color-naming baseline was unexpected. The main problem seems to have been that the 288-trial group was unusually slow in all phases of the experiment.

shape naming. Such a continuum view can handle the results of Experiment 1 as well as other results in the literature (e.g., Glaser & Glaser, 1982).

According to the *continuum of automaticity* view, then, the relatively short periods of practice with shape naming in Experiment 1 could not compete with the lifetime of practice color naming has accrued. The relative speed account makes virtually the same claim, simply omitting any use of the word *automatic* and talking in terms of the relative speed with which each dimension of the compound stimulus is processed. Indeed, it seems to us that these two views coincide very closely at this point. Each amounts to a *differential practice hypothesis* of the sort offered at the end of the last century (cf. Bryan & Harter, 1899) and investigated at the beginning of this century in nonconflict situations (e.g., Lund, 1927). The only difference is in whether the effects are mediated by speed or automaticity of processing. To put such hypotheses to the test, more training in shape naming must be undertaken.

Experiment 2

The principal aim of this experiment was to provide considerably more training in naming shapes. This training should speed up the process of naming shapes with little change in the rate of naming colors, thus permitting a test of the relative speed prediction. By the same token, this training should move the naming of shapes further along the continuum of automaticity without affecting color naming, thereby allowing a test of this view as well. To accomplish this goal, there were 5 days of shape-naming practice. Interference and facilitation were evaluated on the first and last days of training and again about 3 months later (without further training).

According to the relative speed hypothesis, as the distribution of shape-naming times approaches that of color-naming times, the presence of the shapes in Phase 3 should start to affect color naming. Whether a corresponding decline should begin to emerge in the interference and facilitation associated with shape naming (Phase 4) is less clear. Intuition suggests that there should be such a trade-off as training progresses, but this is not required by the relative speed account. Regardless, the 3-month delayed test should permit evaluation of how stable and lasting the effects of practice are in this situation.

The advantage of the relative speed hypothesis is that performance is related directly to an observable performance variable. The continuum of automaticity view does not rely on processing speed as an index, so it is harder to know whether automaticity is being manipulated sufficiently. Still, if 5 days of practice is enough to cause some automatization of shape naming, then incongruent shapes should begin to interfere with color naming. There is nothing about the continuum of automaticity hypothesis that requires interference in the two tasks to trade off over practice; it would be quite reasonable for colors to continue to interfere with shape naming as shapes begin to interfere with color naming.

Method

Subjects. Four volunteers from the same pool as Experiment 1 were paid $5 per session for five 1-hr sessions. Each subject was tested individually, with two subjects assigned at random to each order of Phases 3 and 4. As in Experiment 1, there was no effect of phase order.

Apparatus and stimuli. These were identical to Experiment 1.

Procedure. There were five training sessions spread over 5 days. The first day was carried out exactly as in Experiment 1, using 288 shape-naming practice trials (Phase 2) plus the other three phases. Days 2, 3, and 4 consisted of only shape-naming practice, and 576 practice trials were completed each day, with each shape named equally often. On the final day, everything was the same as Day 1 except that Phase 1 was omitted. (Because the program for baseline color naming had used the four shapes prior to their being assigned names, it would constitute an interference sequence on Day 5.) Overall, there were 2,304 shape-naming trials by Day 5. All other procedures were identical to those in Experiment 1. Subjects were given a chance to relax after every 96 trials on Days 2–4 to reduce fatigue and tedium.

Although not originally planned, a delayed test would reveal how lasting the training effects had been (cf. Kolers, 1976). Consequently, all 4 subjects were contacted again between 3 and 3½ months later to return for a final test. Because our interest was in the persistence of original practice, no practice was given in this delayed session. Only Phases 3 and 4 were carried out, precisely as previously. Subjects were paid $5 each for returning for this final session.

Results and Discussion

As in Experiment 1, error rate was low in all conditions of the experiment. The error rates for shape-naming practice displayed in Figure 1 demonstrated that subjects did not gain speed over days at the expense of accuracy. The error rates for Phases 3 and 4 are shown in Table 2. In Phase 3 (naming colors of colored shapes), there was some suggestion of increased interference between Days 1 and 5 and especially after 3 months, but the error data were far from reliable. In Phase 4 (naming shapes of colored shapes), consistent interference was evident. There was little evidence of facilitation anywhere in the accuracy data, and no suggestion of a speed–accuracy tradeoff in the analyses.

The primary data to be discussed are the mean correct naming latencies in each condition. Figure 1 shows the drop

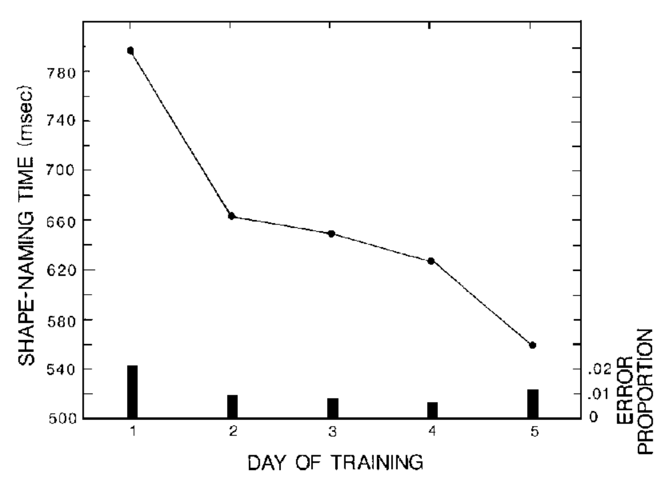

Figure 1. Experiment 2: The decline in mean shape-naming time as a function of days of practice (Days 1–5). (Also shown at the bottom of the figure is error rate on each day.)

Table 2
Experiment 2: Mean Response Times (in Milliseconds) and Proportions of Errors (E) for Days 1 and 5 of Training, and for 3 Months Later

Session	Phase 1: Baseline colors	Phase 2: Baseline shapes	Phase 3: Naming colors on shapes			Phase 4: Naming shapes in color		
			CG	CT	IN	CG	CT	IN
Day 1								
RT	678	799	625	614	640	669	733	805
E	.020	.021	.021	.000	.021	.021	.042	.104
Day 5								
RT		560	605	614	699	630	661	733
E		.011	.021	.021	.042	.000	.000	.104
3 months								
RT			596	620	722	583	634	692
E			.000	.021	.146	.000	.000	.062

Note. The three conditions in Phases 3 and 4 are congruent (CG), control (CT), and incongruent (IN).

in mean shape-naming time over the five days of practice.[2] This decline was highly significant, $F(4, 12) = 42.14$, $p < .001$, $MS_e = 566.16$. Mean shape-naming time dropped 239 ms from 799 ms to 560 ms. This final value was well under the Day 1 color-naming baseline (678 ms) and the Day 1 and Day 5 color-naming control conditions from Phase 3 (both 614 ms). Note, however, that shape-naming time was longer on Day 5 in the control condition of Phase 4 than in the baseline shape-naming condition (661 ms versus 560 ms). Apparently, processing on control trials is affected by the proximity of congruent and incongruent trials. Still, the important point is that the distribution of shape-naming times had moved much closer to that of color-naming times after five days of training.

Table 2 displays the naming time data for Days 1 and 5, as well as for the 3-month delayed test. To examine the effects of training on each naming task, 2 × 3 ANOVAS were carried out separately on the color-naming (Phase 3) and shape-naming (Phase 4) data. The variables were day of practice (1 and 5) and condition (congruent, control, and incongruent). First, consider shape naming in Phase 4. As expected, there was a significant effect of days of practice, $F(1, 3) = 9.93$, $p < .05$, $MS_e = 2{,}279.15$, showing that overall shape-naming times decreased in Phase 4 as in Phase 2. As in Experiment 1, shape-naming time was strongly affected by the presence of colors, as shown by the significant effect of condition, $F(2, 6) = 34.13$, $p < .001$, $MS_e = 849.49$. On both Days 1 and 5, interference was 72 ms; facilitation was 64 ms on Day 1 and 31 ms on Day 5. Only the 31 ms of facilitation was not significant at the .05 level using Dunnett's test. That the interference and facilitation effects were consistent from the beginning to the end of practice is indicated by the absence of any interaction between day of training and condition, $F(2, 6) = 0.29$, $p > .10$, $MS_e = 2{,}461.57$.

The color naming of Phase 3 reveals a quite different picture. Although there was a significant main effect of condition, $F(2, 6) = 7.37$, $p < .05$, $MS_e = 1{,}079.17$, this was coupled with a significant interaction of days of practice and condition, $F(2, 6) = 6.89$, $p < .05$, $MS_e = 485.88$. There was no suggestion of any facilitation from compatible shapes and colors on either day, with difference scores of 11 ms in the wrong direction on Day 1 and 9 ms in the right direction on Day 5. However, interference grew from a nonsignificant 26 ms on Day 1 to a significant 85 ms on Day 5, as shown by Dunnett's test. Thus, interference emerged in color naming with practice on shape naming. The nonsignificant effect of days of practice, $F(1, 3) = 0.12$, $p > .10$, $MS_e = 8{,}154.39$, merely reflected the fact that color naming was not practiced, and hence did not change much over days.

After 3 months without further training, the Day 5 pattern persisted. Both Phases 3 and 4 showed significant effects of compatibility between the two dimensions of the stimuli: for shape-naming, $F(2, 6) = 26.94$, $p < .001$, $MS_e = 392.14$; for color naming, $F(2, 6) = 7.34$, $p < .05$, $MS_e = 2{,}284.53$. In shape naming, there were 51 ms of facilitation and 58 ms of interference; in color naming, there were 24 ms of facilitation and 102 ms of interference. Only the 24 ms of facilitation in color naming was nonsignificant by Dunnett's test. As Kolers (1976) has shown with reading transformed text, these sorts of training effects are remarkably long lasting.

The central finding in Experiment 2 is the emergence of interference in color naming that is due to practice in naming shapes. Although the 2 days of training in Experiment 1 were not enough to produce interference in color naming, the 5 days in Experiment 2 were. Each of the two dimensions of the compound stimuli interfered with processing of the other dimension. Both the continuum of automaticity and the relative speed of processing accounts predict that interference and facilitation should begin to appear in both directions after sufficient practice. At this level of analysis, then, both hypotheses correctly predicted the observed pattern. But one aspect of the data seems inconsistent with these views. On Day 5, interference was 84 ms in color naming and 72 ms in shape naming. These values are very close; if anything, the newly learned dimension (shape names) seems to have caused more interference with the familiar dimension (color names) than vice versa. The same was true 3 months later, where color naming showed 102 ms of interference and shape naming showed 58 ms of interference.

What happened in the 5 days of practicing shape naming? According to the relative speed of processing account, shape name processing speeded up, moving closer to color name

[2] On Day 3, the data for 192 shape-naming practice trials were lost for 1 subject. Mean shape-naming time for that day was estimated from the remaining 384 trials.

processing. Although Figure 1 suggests that shape processing times had not yet asymptoted by Day 5, they had moved very close to color processing times. What is interesting is that the amounts of interference were almost equal in the two directions on Day 5—this is predicted to occur when the distributions overlap—but that interference in *both* directions after 5 days (85 ms for naming colors and 72 ms for naming shapes) was so similar to that on Day 1 in Phase 4 shape naming (72 ms). Intuitively, it might seem most reasonable that total interference in both directions on Day 5 should equal that in shape naming alone on Day 1; instead, all three values are equal. There is no trade-off as the distributions of processing times move together.

According to the continuum of automaticity view, shape naming became more automatic, moving closer to color naming on that conceptual dimension. While intuition and existing theory (e.g., Shiffrin & Schneider, 1977) might suggest that 5 hr of practice on shape naming ought not to compensate for a lifetime of practice in recognizing and naming colors, apparently it did. The two dimensions interfered with each other almost equally after 5 days of shape-naming practice, and this effect persisted after 3 months without further training. Because there is no asymmetry, we must assume under the continuum of automaticity hypothesis that the two processes are roughly equally automatic.

Experiment 3

What would happen if training was carried considerably farther, to speed up shape naming even more and to increase its automaticity further? Under the relative speed of processing account, gradually the color should stop interfering with naming the shape and the shape should begin to cause more interference with naming the color. This is predicted because the distribution of shape-naming times should progressively become faster than the distribution of color-naming times, and the claim is that faster processes interfere more frequently with slower ones than vice versa.

Under the continuum of automaticity account, as shape naming becomes more automatic with practice, shapes should interfere more with color naming. However, it is unclear whether the interference in shape naming caused by colors should change. Increasing the automaticity of one process does not necessarily have implications for the automaticity of any other process. Thus, the continuum of automaticity view does not require a trade-off, although versions of this view could be suggested that could handle such a trade-off. Note that this is a problem with the continuum of automaticity view: Although the relative speed of processing explanation is tied directly to observable response times, the automaticity view is not, at least without modification of the task.

Experiment 3 undertook to increase dramatically the extent of practice at shape naming with the goal of determining whether a shift in the pattern of interference would develop. Consequently, subjects did 20 days of practice naming shapes, a total of 10,656 practice trials (2,664 trials per shape). They were tested on the interference and facilitation phases only on Days 1, 5, and 20, with Day 5 included to permit comparability of the data to Experiment 2. Otherwise, every day consisted of practicing shape names only.

Method

Subjects. There were 4 subjects: 3 undergraduate students and 1 graduate student at the Scarborough Campus of the University of Toronto.[3] Two had Phase 3 before Phase 4 on Days 1, 5, and 20 and 2 had these two phases in the opposite order. As in the previous experiments, order of phases did not affect the data. Each subject was paid $100 for completing the 20 sessions of the experiment.

Apparatus and stimuli. These were identical to Experiments 1 and 2, with the change in Phase 1 noted in the Procedure section.

Procedure. Although largely the same scheme was used here as in Experiment 2, two modifications were made. First, the color-naming baseline (Phase 1) procedure was modified so that only squares—not the to-be-learned shapes—were seen in color. In this way, Phase 1 could be carried out on Days 5 and 20 (in addition to Day 1) without effects attributable to training, and a daily baseline could be obtained. Otherwise, the four Phases all were carried out on Days 1, 5, and 20 precisely as in Experiment 2. On Days 2 through 4 and 6 through 19, only Phase 2 was run, with 576 trials a day.

The second modification permitted the procedure to be carried out without need for an experimenter. Previously, an experimenter had determined the accuracy of every trial and recorded its value after each trial. This was exceedingly boring and time consuming in a 5-day study; it would have been unbearable in a 20-day study. Thus, the programs were rewritten so that after each oral response (on which the response latency was based), the computer displayed what the response should have been, and the subject pressed one of two buttons to indicate whether the correct response had been made. Subjects were also advised to press the error button if their response failed to trip the voice key (a relatively rare occurrence).

Subjects were given keys to the laboratory and were told to do a daily session every Monday through Friday for 4 weeks. They were asked to keep to the same time schedule as much as possible each day. For 2 subjects, a holiday intervened; they tested themselves twice the preceding day, once early and once late in the day. Another subject had to do two sessions one day because he missed the preceding day.

Results and Discussion

We will survey the error data first before turning to the latency data of primary interest. Over the 20 sessions of shape name training (Phase 2), the 4 subjects had mean error rates of .004, .029, .011, and .026. On no daily training session did the mean error rate, collapsed over the four subjects, exceed .024 (see the bottom of Figure 2). Thus, subjects were highly accurate in labeling shapes throughout training. As Table 3 shows, the error rates for baseline color naming (Phase 1) also were low. The only error rates that stand out are the incongruent trial values in Phase 4 on Days 1, 5, and 20, and in Phase 3 on Days 5 and 20. These suggest interference in these conditions. As we will discuss shortly, they are quite consistent with the latency data. There was no evidence at all of any speed–accuracy trade-off in the data.

Figure 2 displays the naming latencies for shapes over the 20 days of practice. Subjects sped up considerably with practice, from a starting value on Day 1 of 628 ms to a final value

[3] Three additional subjects took part in this experiment, but their data are not included. For one, it was discovered in the first week that his Day 1 data had not been recorded. Because he would have no baselines, he was excused from further testing. The shape-naming times of the other 2 did not improve much with practice, so they were replaced.

Table 3
Experiment 3: Mean Response Times (in Milliseconds) and Mean Proportions of Errors (E) for Each of the Four Phases on Days 1, 5, and 20 of Training

Session	Phase 1: Baseline colors	Phase 2: Baseline shapes	Phase 3: Naming colors on shapes			Phase 4: Naming shapes in color		
			CG	CT	IN	CG	CT	IN
Day 1								
RT	512	628	506	511	518	588	622	703
E	.025	.021	.010	.021	.010	.000	.031	.062
Day 5								
RT	476	531	520	520	590	548	579	680
E	.000	.015	.010	.010	.031	.010	.021	.042
Day 20								
RT	425	421	499	483	594	459	466	486
E	.019	.016	.000	.042	.114	.010	.042	.062

Note. The three conditions in Phases 3 and 4 are congruent (CG), control (CT), and incongruent (IN).

on Day 20 of 421 ms, a gain of 207 ms. This improvement was highly significant, $F(19, 57) = 15.19$, $p < .001$, $MS_e = 751.12$.

Table 3 displays the latency data for the critical days: Days 1, 5, and 20. Note first that the subjects in Experiment 3 were considerably faster in all phases than were those in Experiment 2. This may well stem from the increased emphasis on speed in their instructions, aimed at maximizing their improvement in shape naming with training.

In keeping with previous analyses, we will evaluate the color-naming and shape-naming data separately, using 3 × 3 (day by condition) repeated-measures ANOVA. Consider first the color-naming data of Phase 3. The effect of day of training was nonsignificant, $F < 1$, whereas the effect of condition was significant, $F(2, 6) = 8.72$, $p < .02$, $MS_e = 1{,}719.71$. The interaction of day with condition was marginally significant, $F(4, 12) = 2.55$, $p = .09$, $MS_e = 1{,}238.49$. No facilitation appeared on any of the three days. Like Experiment 2, there was virtually no interference on Day 1 but strong interference appeared by Day 5. Furthermore, this interference persisted through to Day 20, actually growing somewhat. As shape-naming practice progressed, the potential of the shapes to cause interference in color naming increased.

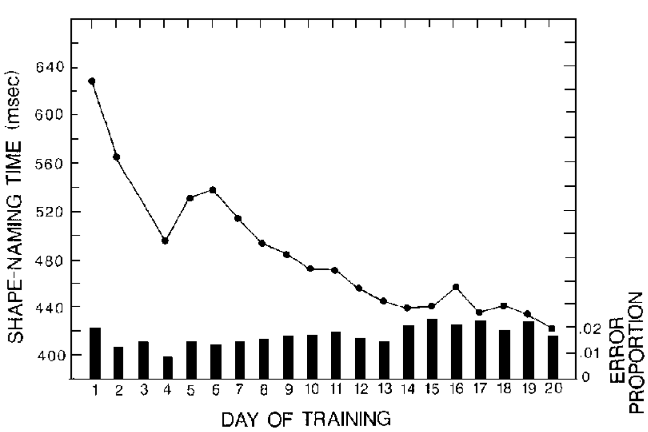

Figure 2. Experiment 3: The decline in mean shape-naming time as a function of days of practice (Days 1–20). (Also shown at the bottom of the figure is error rate on each day. Note that Days 6, 11, and 16 followed weekends, accounting for the slight increase in naming time on these days.)

Now consider the shape-naming data of Phase 4. Both the main effect of day of training, $F(2, 6) = 19.15$, $p < .01$, $MS_e = 4{,}881.41$, and the main effect of condition, $F(2, 6) = 17.25$, $p < .01$, $MS_e = 1{,}556.55$, were significant. The interaction was marginally significant, $F(4, 12) = 2.67$, $p = .08$, $MS_e = 1{,}264.42$. Days 1 and 5 replicate the pattern in Experiment 2: There was strong interference coupled with modest facilitation. The striking result occurred on Day 20, where both the facilitation and the interference diminished sharply. For the first time, incompatible colors did not interfere significantly with shape naming.

The 20-day data are the most interesting and novel data of Experiment 3. Notice first that by Day 20 the baseline color-naming (Phase 1) and shape-naming (Phase 2) times were almost identical. Furthermore, the control conditions for the two-dimensional stimuli (Phases 3 and 4) were also very similar. Although incompatible shapes still interfered with color naming (Phase 3), incompatible colors no longer interfered with shape naming (Phase 4). Apparently, 20 days of training was sufficient to turn the tables. On Day 1, familiar color names interfered with and facilitated naming unfamiliar shapes, but not vice versa. On Day 20, heavily practiced shape names interfered with naming familiar colors, but not vice versa.

Given the equivalence of baseline processing times and control processing times for colors and shapes on Day 20, the relative speed of processing account would predict equivalent facilitation and interference in both directions. Yet there is clearly a strong asymmetry, with interference only shown in naming colors. These data contradict the relative speed of processing hypothesis.

The continuum of automaticity explanation fares considerably better, although this may be only because we have no directly observable index of extent of automaticity. Assuming that 20 days of practice has pushed shape naming well beyond color naming in terms of degree of automaticity, the observed pattern is just what this view predicts. If the degree of automaticity of two processes is sharply imbalanced, the more automatic dimension should selectively interfere with the less automatic dimension. Apparently, the difference in automaticity is more telling than the absolute levels of automaticity.

Anecdotally, subjects certainly claim that shape naming has become dominant in their repertoire. One subject reported dreaming about the shapes, and another said that the shapes began to take on the color of the names assigned them even on training days, when they appeared in white. A third subject reported seeing a lake whose outline resembled that of one of the shapes and immediately thinking of that shape's name. Shape naming may not have become faster than color naming in 20 days, but it certainly became dominant and quite effortless.

General Discussion

Over three experiments, we have observed a progression in the pattern of interference and facilitation based on the extent of practice subjects had with one dimension of a two-dimensional task. Initially, subjects were trained only briefly—1 or 2 days—on giving unfamiliar shapes new color names. When asked to name the shapes in color, incongruent colors interfered with shape naming and congruent colors facilitated shape naming. However, the presence of shapes had no effect on color naming. Put simply, there was a strong asymmetry.

Five days later in training, interference became symmetrical, occurring in both directions to the same extent that it occurred earlier in only one direction. Although facilitation from congruent stimuli persisted when naming shapes, it did not appear when naming colors. This pattern carried over to a test some 3 months later without any further practice on shape naming. Finally, after 20 days of shape-naming practice, the original asymmetry was reversed. At this point, the presence of colors no longer had any influence on the naming of shapes. However, an incongruent shape now interfered strongly with color naming, although without any corresponding facilitation.

Consider the interference data from the standpoint of the relative speed of processing view. The data from early in training seem quite consistent with the predictions from this explanation. Basically, the slower process—shape naming—is affected by the faster process—color naming—but not vice versa. But as training progresses, the relative speed hypothesis begins to stumble. After 5 days of shape-naming practice, shape naming has become much faster than it was to begin with, although it is still not as fast as color naming. Although interference in shape naming is unchanged, there is now interference in color naming of equivalent magnitude. Apparently, there is no trade-off in interference as the distributions of shape-naming and color-naming times come closer together.

The fatal blow for the speed of processing account is delivered by the Day 20 data. By this point in training, the two dimensions are processed at roughly equal rates. The straightforward prediction from the relative speed hypothesis would be for equivalent interference in the two tasks, but this point had already passed by Day 5. Instead, the asymmetry of early in practice now is reversed, and only color naming shows interference from incongruent shapes. Such a finding is completely at odds with the relative speed of processing explanation.

The continuum of automaticity view can accommodate the interference results quite comfortably. On Day 1, color naming was more automatic than shape naming. By Day 5 they were roughly equally automatic, and by Day 20 shape naming was considerably more automatic than color naming. Such an argument is completely in keeping with the idea of a continuum of automaticity and even opens up the possibility that a potent degree of automaticity may be possible in some tasks with fairly minimal training.

Of course, we hasten to note that the continuum of automaticity view currently has the edge by virtue of not being tied to any "thermometer" of automaticity in our experiment. Nevertheless, our point is simply that the continuum view is a *possible* explanation whereas the relative speed view does not seem to be viable. One direction for future research is to develop and test indices of automaticity that are independent of the response latency data in the task of primary interest. An obvious candidate is the use of the secondary task methodology (cf. Shiffrin, in press), but there may be other techniques that would be suitable as well. Also worth pursuing is the extent to which degree of automaticity of processing some dimension is task dependent.

Until now, we have avoided discussing facilitation in the Stroop tasks. In our work with transformed color words (Dunbar & MacLeod, 1984), we found that facilitation and interference often came decoupled. It is interesting that the only place in the present series of experiments where we observed facilitation was when it co-occurred with interference in shape naming on Days 1 and 5. That is, congruent colors facilitated shape naming before shape naming became more automatic than color naming. Notice that facilitation of shape naming disappeared along with interference on Day 20 and, more important, we never observed significant facilitation for color naming at any point in training.

The facilitation results initially seem at odds with the relative speed hypothesis. Under this view, larger interference should predict larger facilitation, yet our largest interference was in color naming on Day 20 of Experiment 3, and there was no corresponding facilitation. Although in post hoc fashion, the continuum of automaticity view can accommodate these findings. It is as if the newly learned shapes were sufficiently automated to interfere with color naming, but not to facilitate it. Facilitation may require much greater automaticity than does interference, which would explain its less frequent appearance here and in other studies (e.g., Dunbar & MacLeod, 1984). Alternatively, facilitation may rely on a wholly different mechanism than does interference.

There is another explanation of facilitation that is more methodological. Because facilitation appears only when the control condition is quite slow (i.e., before practice on unfamiliar dimensions), it may be that we are dealing with a measurement problem, not a conceptual one. If the control condition is fast, there may be no room for the congruent condition to be faster. Such a view would predict that facilitation should be less likely as practice proceeds, in contrast to the continuum of automaticity prediction. This represents another direction for future work on the Stroop effect: to determine the relation between interference and facilitation.

Studies manipulating the ratios of different trial types might be profitably pursued in this regard.

In the future, further training experiments should be undertaken in which training on each of the two dimensions is manipulated independently, and in which some extra task measure of the extent of automaticity of each dimension is taken. For the present, however, we believe that the bulk of the evidence favors a continuum of automaticity account over a relative speed of processing account.

It certainly is apparent that speed of processing alone cannot predict patterns of interference and facilitation in the Stroop task. As Bryan and Harter said in 1889, "No reaction time test will surely show whether a given individual has or has not effective speed in his work . . . .The rate at which one makes practical headway depends partly upon the rate of mental and nervous processes involved; but far more upon how much is included in each process" (pp. 374–375). This suggests an interesting possibility for what happens as automaticity increases: Processes become more comprehensive or integrated. The superficially simple Stroop task may not, in the end, be simple, but its complexities may help us to develop a deeper understanding of our basic attentional processes.

References

Brown, W. (1915). Practice in associating color-names with colors. *Psychological Review, 22,* 45–55.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345–375.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65.

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 and Performance, 10,* 622–639.

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition, 1,* 106–120.

Glaser, W. R., & Dolt, M. O. (1977). A functional model to localize the conflict underlying the Stroop phenomenon. *Psychological Research, 39,* 287–310.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 875–894.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

James, W. (1890). *Principles of psychology.* New York: Holt.

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 and Performance, 9,* 497–509.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 554–565.

Ligon, E. M. (1932). A genetic study of color naming and word reading. *American Journal of Psychology, 44,* 103–122.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367–386.

Lund, F. H. (1927). The role of practice in the speed of association. *Journal of Experimental Psychology, 10,* 424–433.

Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology, 25,* 387–397.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Pritchatt, D. (1968). An investigation into some of the underlying associative verbal processes of the Stroop colour effect. *Quarterly Journal of Experimental Psychology, 20,* 351–359.

Shiffrin, R. M. (in press). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens's handbook of experimental psychology* (2nd ed.). New York: Wiley.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology, 31,* 121–132.

Stroop, J. R. (1935a). The basis of Ligon's theory. *American Journal of Psychology, 47,* 499–504.

Stroop, J. R. (1935b). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Vanderplass, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology, 57,* 147–154.

Received February 14, 1985
Revision received January 26, 1987
Accepted February 25, 1987 ■