Journal of Experimental Psychology:
Human Perception and Performance
2005, Vol. 31, No. 3, 438–452

Copyright 2005 by the American Psychological Association
0096-1523/05/$12.00 DOI: 10.1037/0096-1523.31.3.438

Reduced Stroop Interference for Opponent Colors May Be Due to Input Factors: Evidence From Individual Differences and a Neural Network Simulation

Bruno Laeng, Torstein Låg, and Tim Brennen
University of Tromsø

Sensory or input factors can influence the strength of interference in the classic Stroop color–word task. Specifically, in a single-trial computerized version of the Stroop task, when color–word pairs were incongruent, opponent color pairs (e.g., the word *BLUE* in yellow) showed reduced Stroop interference compared with nonopponent color pairs (e.g., *BLUE* in red). In addition, participants' color discrimination ability was measured by standard color vision tests (i.e., Farnsworth-Munsell 100-Hue Test and Ishihara plates). Error rates in the Farnsworth-Munsell test correlated positively with the amount of Stroop interference. Neural network simulations (variants of J. D. Cohen, K. Dunbar, & J. L. McClelland's, 1990, model) showed that only a distributed trichromatic input layer was able to simulate these findings. Thus, sensory input from the color system needs to be incorporated into current accounts of the Stroop effect.

Keywords: Stroop effect, opponent colors, color vision, individual differences, neural networks

The Stroop effect, in either its classic form (i.e., naming the color of colored words; Stroop, 1935) or its less colorful variations (e.g., naming drawings overlapping words), has been central in cognitive psychology as well as of great interest for clinical psychologists, psychiatrists, and neuroscientists (cf. Brown, Gore, & Carr, 2002; MacLeod, 1991). The continuing and renewed interest in what outside of academia is considered a rather amusing and bizarre verbal teaser may be due to the belief that Stroop effects are helpful in understanding attention, perception, and reading, as well as the cognitive and neural mechanisms behind inhibitory effects, psychological mechanisms like interference, controlled versus automatic processing, and so forth.

In the present study, we examined the effect of input factors on Stroop interference. Hock and Egeth (1970) might have been the first to propose a perceptual-encoding account where one of the sources of interference is hypothesized to occur at an early selection stage, that is, the perceptual-encoding stage of the ink-color information. More recent models (e.g., Kornblum, Stevens, Whipple, & Requin, 1998; Melara & Algom, 2003; Zhang & Kornblum, 1998; Zhang, Zhang, & Kornblum, 1998) have clearly stressed the role of perceptual processes in the Stroop color-naming effect. Some have looked at the relations between various modes of stimulus input and response output (e.g., Lu & Proctor, 2001; Virzi & Egeth, 1985).

As pointed out by MacLeod (1991), there seems to be a widespread belief that it is the processing subsequent to encoding that is most important in the Stroop effect. Perhaps because of this belief, most computational models of the Stroop effect have not given specific relevance to input factors. Several of the available and most advanced network simulations (e.g., Cohen et al., 1990; Kello, Plaut, & MacWhinney, 2000; Phaf, Van der Hejden, & Hudson, 1990; Roelofs, 2003; Zhang et al., 1998) have not explored specific effects of changes at the input layer. In particular, no attempt has been made to simulate basic properties of human color receptors and early color-processing stages, like trichromacy—that is, the organization of input by long (L), medium (M), and short (S) cone types—or color opponency—that is, $L - M$ and $S - (L + M)$ mechanisms. Therefore, the influence of such variables on the behavior of these networks, as well as their ability to reproduce human data, is still not known. Yet, we surmise that some of the known effects of associative strength or semantic distance of color words from the to-be-named colors (e.g., Klein, 1964; Klopfer, 1996; Sichel & Chandler, 1969) may be accounted for, at least in part, by the particular way in which the input layer is organized.

A simple way to study the role of perceptual processes on the Stroop effect could be to investigate individual differences at the input, or sensory, level of color processing. Individual differences in sensory discriminations may also be conceptualized as changing the landscape of color space. A study by Klopfer (1996) has shown that color naming becomes increasingly difficult when colors and words refer to closer-by regions of color space. One interpretation of this finding is that distance in color space and degree of overlap

Bruno Laeng, Torstein Låg, and Tim Brennen, Department of Psychology, University of Tromsø, Tromsø, Norway.

We thank Ernest Dalrymple-Alford, Bradley Gibson, Dale Kopfler, Robert Proctor, and Carlo Umiltà for their insightful comments on drafts of this article. For their help in data collection and in discussions of the project, we are also very grateful to Yngvild Arnesen, Guro Brandvik, Camilla Breivik, Thomas Espeseth, Toril Fossmo, Knut Hansson, Olav Kandel, Kjetil Karstad, Cecilie Larsen, Kathrine Rasch, Eline Røed, Synnøve Simonsen, Jill Terres, Oddrun Vermeer, and Marianne Voldnes.

Correspondence concerning this article should be addressed to Bruno Laeng, Department of Psychology, University of Tromsø, Huginbakken 32, N-9037 Tromsø, Norway. E-mail: bruno@psyk.uit.no

between such regions determine the extent of processing (e.g., selective attention) needed to discriminate between the two. Near neighbors in color space would then be associated with increased performance time. Thus, assuming that the color and the word (naming a different color) activate different regions in color space, these regions might differ among individuals, both in extent and overlap. For example, in a Stroop trial where the color is green and the word is *blue,* the two activated regions of color space would be further apart in an individual with good blue–green discrimination ability than in another individual with less good blue–green discrimination. Consequently, the word *blue* would interfere to a lesser degree in individuals who are able to discriminate well between blue and green than in individuals who make poor discriminations between these two colors.

At its inception, the investigation of the interference of words' meanings on object properties' naming was mainly motivated by an interest in individual psychological differences (cf. Jensen & Rohwer, 1966). Indeed, individual differences in the degree of Stroop interference are known to be among the most reliable or stable measures. Moreover, despite individual variation in the size of the effect, the relative differences of scores of congruently and incongruently colored words (as well as those of control conditions) maintain an impressive regularity at the group level, and the same rank order of magnitude can be observed for nearly all participants (e.g., Jensen & Rohwer, 1966, did not find a single exception in a study of over 400 participants).

Study Hypotheses

1. If some of the interference in the color–word Stroop effect originates at the perceptual level, then differences in color perception skills among individuals should have an impact on the degree of Stroop interference. To our knowledge, although there are several studies that have addressed the influence of individual differences on the Stroop effect, for example, Kane and Engle's (2003) study of individual differences in working memory, no studies have systematically addressed the influence of individual differences in color vision on the Stroop effect.

2. If the similarity space among colors influences the amount of interference between the color and the word, then we would expect to find that color–word pairs with the largest distance in color space (i.e., opponent colors, like blue and *yellow*) would show a smaller Stroop effect (i.e., difference in response times [RTs] between incongruent and congruent trials) compared with color–word pairs with an intermediate distance in color space (i.e., nonopponent color–name pairs, like blue and *red*). Effects of similarity between the ink color and the color denoted by the word have been previously observed in studies that manipulated the distance between the colors (e.g., Klein, 1964; Klopfer, 1996). In addition, we would expect that individuals who have difficulty with a specific region of color space should show more susceptibility to Stroop interference when the color–word pairs include the poorly discriminated color.

3. Finally, the trichromatic neural coding of the human retina is undoubtedly one of the basic factors underlying the specifically human organization of color space. If so, implementing the trichromatic neural coding at the input layer of an artificial network might be sufficient to simulate the reduced Stroop interference of opponent color–word pairs. We explored the role of input factors in a variant of Cohen et al.'s (1990) network model of the classic Stroop effect.

Experiment 1

In this experimental study with human participants, we used a single-trial computerized version of the classic color–word Stroop task (cf. Salo et al., 2001), with color and word pairs forming either congruent or incongruent matches. Two groups of participants responded by pressing keys labeled with appropriate color patches (each group with different orderings of the labels), and a third group gave verbal responses. All combinations of color–name pairs between the two basic pairs of opponent colors (i.e., blue, yellow, red, and green) were used.

First, we expected to find that color–word pairs with opponent colors would reduce the strength of Stroop interference compared with nonopponent color–name pairs. This finding would replicate aspects of the similarity effects observed in previous studies (e.g., Klein, 1964; Klopfer, 1996). Second, we expected to find that participants' differences in their color discrimination abilities, as measured by the Farnsworth-Munsell 100-Hue (FMH) Test (Farnsworth, 1957) and the Ishihara plates (Ishihara, 1951), would influence the strength of interference in the classic Stroop color–word task. Specifically, error rates in the color tests should correlate positively with the strength of the Stroop interference.

Method

Participants. One hundred eighty-two students (mean age = 23, *SD* = 6) at the University of Tromsø (104 female and 78 male participants) volunteered for a color–word Stroop experiment. Among these participants, 138 (72 female and 66 male participants) also received the color vision tests.

Stimuli. The words *BLÅ, GRØNN, RØD,* and *GUL* (48-point Geneva, Switzerland, font) were presented at the center of the computer screen over a white background. Henceforth, these will be referred to as *BLUE, GREEN, RED,* and *YELLOW.* Each word could be colored in blue (red, green, blue [RGB] coordinates = 0, 0, 255), green (RGB coordinates = 0, 255, 0), red (RGB coordinates = 255, 0, 0), or yellow (RGB coordinates = 255, 255, 0).

Apparatus and procedure. The *G, H, J,* and *K* keys were relabeled with focal blue, green, red, and yellow color patches. For one group of participants (N = 111, women = 68), the color patches were in the following sequence (from left to right): red, blue, green, and yellow (the RBGY set). For another group (N = 42, women = 21), the color patches were in the following sequence: green, red, yellow, and blue (the GRYB set). This second group of participants was included as a control condition to establish that the hypothesized differences in interference between color–word pairs were not due to the specific ordering of the color-labeled response keys. Participants were requested to press with the index finger of their preferred hand, as quickly and as accurately as possible, the key with the color patch corresponding to the color of the word. A keypress terminated the stimulus presentation while the PC recorded the correctness of the keypress and the RT from the onset of the word to the key response. The third group of participants (N = 29; women = 15) responded by naming aloud the word's color, which activated a microphone switch that terminated the stimulus presentation and the RT measurement. For all groups of participants, trials were self-paced with participants pressing the space key to begin a trial; participants in the manual response groups were also

instructed to use the index finger of their preferred hand to press the space key (so that all of the color patches would be visible before responding).

The testing of color vision consisted of (a) the FMH test and (b) the Ishihara plates. The FMH test is a simple and effective method for testing color discrimination. It consists of four sets of plastic caps in which the colors are mounted. There are 85 moveable caps of the same brightness representing hues along the complete human circular color space (equally in all strengths from neutral to high purity). The task is to rearrange the caps, from an initial random arrangement, according to color similarity between two fixed reference caps. The error score for a cap is calculated as the sum of the differences between the caps adjacent to it; hence the minimum score of 2 is different from zero. Mantere, Parkkinen, Mäntyjärvi, and Jasskelainen (1995) have calculated the correlation matrix of the cone responses of the 85 FMH test caps and have determined eigenvectors of the matrix; this showed that the first three eigenvectors can be interpreted as the opponent color signals and the nonopponent color signal. Thus, observers with normal color vision can determine the color of a cap by using two opponent color signals; for color-blind persons (dichromats), one or the other opponent signal is defective, and errors occur during the test. The test was administered by placing the caps on a table under a halogen lamp (about 6500° Kelvin illumination). Scoring of errors was performed according to the standardized FMH test scoring system.

The Ishihara test consists of 14 pseudoisochromatic plates made of colored circles of various sizes, representing double-digit numbers inscribed inside a circle. The Ishihara test is typically used clinically for testing color perception. In the task, participants must simply name each of the numbers (allowing 2 s for a response). Omissions and incorrect responses are recorded on a score sheet by the experimenter. Incorrect responses to 4 or fewer plates indicate normal vision, whereas incorrect responses to 5 or more plates indicate defective red–green vision; however, high error scores in the test do not classify the type of red–green defect or the amount of defect.

There were a total of 96 trials in the Stroop task; 48 trials showed congruent color–word pairs (e.g., *BLUE* in blue), 24 trials showed incongruent opponent color–word pairs (e.g., *BLUE* in yellow), and 24 trials showed incongruent nonopponent color–word pairs (e.g., *BLUE* in red). The trials' presentation order was completely randomized by use of SuperLab software (Abboud & Sugar, 1990). Half of the 138 participants who completed all tests received the color vision test before the Stroop task, whereas the other half performed the tasks in the reverse order. Sex as a variable was taken into account in the following analyses. Although the presence of sex differences in Stroop interference remains controversial (cf. MacLeod, 1991, p. 184), it is well known that female participants have better color vision than male participants (cf. Pickford, 1951).

Results

The opponent colors reduced Stroop effect. We first calculated descriptive statistics for each participant, obtaining mean RTs for correct responses and error rates for each combination of the variables word color (i.e., blue, green, red, and yellow) and word name (i.e., *BLUE, GREEN, RED,* and *YELLOW*).

Error rates were generally low. Mean error rates (in percentages) were calculated for each participant on the basis of those trials where the word name spelled out the word color (e.g., *BLUE* in blue); this condition was named the congruent color condition (mean error = 2.4%, SD = 3.2). Similarly, mean error rates were calculated for trials that paired opponent colors (i.e., *BLUE* in yellow, *GREEN* in red, *RED* in green, and *YELLOW* in blue); this condition was renamed the opponent color condition (mean error = 3.1%, SD = 3.4). Finally, mean error rates were calculated for trials that paired nonopponent colors (i.e., *BLUE* in green, *BLUE* in red, *GREEN* in blue, *GREEN* in yellow, *RED* in blue, *RED* in yellow, and *YELLOW* in green, *YELLOW* in red); this condition was renamed the nonopponent color condition (mean error = 2.5%, SD = 2.7). A repeated-measures analysis of variance (ANOVA) with condition (congruent, opponent, nonopponent) as the within-subject variable and sex (female, male) and response type (the GRYB set of keys, the RBGY set of keys, microphone) as the between-subjects variables was performed on mean percentage of errors as the dependent variable. This showed no significant effect of condition, $F(2, 352) = 0.8$, $p = .41$, or of any of the other variables and their interactions.

The same averaging procedure described above was applied to RTs to obtain, for each participant, mean RTs in the congruent, opponent, and nonopponent color conditions. RTs from trials on which errors occurred were excluded from analyses of the RTs, and trials with RTs greater than 3 standard deviations from each individual's mean RT for that cell were treated as outliers and excluded from all subsequent analyses (less than 1% of the data in each condition were excluded by this trimming rule).

The first analysis performed on RTs as the dependent variable was a repeated-measures ANOVA with condition (congruent, opponent, nonopponent) as the within-subject variable, and sex (female, male) and response type (the GRYB set of keys, the RBGY set of keys, microphone) as the between-subjects variables. The analysis revealed a main effect of condition, $F(2, 352) = 80.9$, $p = .0001$. As expected (see Figure 1), participants evaluated congruent color–word pairs (mean RT = 758 ms, SD = 141) faster than opponent pairs (RT = 810 ms, SD = 147), whereas nonopponent pairs were evaluated the slowest (mean RT = 848 ms, SD = 168). Fisher's protected least significant difference (PLSD) post hoc tests confirmed that the differences were significant for each combination of conditions (critical difference = 31.4 ms; $-38 <$ mean differences < -90; $.02 > p = .0001$).

There was also a main effect of sex, $F(1, 176) = 7.9$, $p = .005$. Female participants evaluated color–word pairs faster (mean RT = 788 ms, SD = 160) than the male participants (mean RT = 834 ms, SD = 148). Also, sex interacted with condition, $F(2, 352) = 4.1$, $p = .005$. To assess whether the sex difference appeared not only in the absolute speed for different conditions but also in the strength of the Stroop effect, we performed an additional ANOVA with sex (female, male) as the between-subjects variable and Stroop strength (i.e., the difference between the two averaged, incongruent conditions and the congruent condition) as the dependent variable. This confirmed a larger Stroop effect in men (mean RT = 82.1, SD = 61) than in women (mean RT = 61.9, SD = 51), $F(1, 180) = 5.9$, $p = .02$.

In addition, there was a main effect of response type, $F(2, 176) = 6.3$, $p = .002$. Participants evaluated color–word pairs the slowest with verbal responses (mean RT = 852 ms, SD = 201) and the fastest when matching keypresses to color patches with opponent colors adjacent to one another, that is, the GRYB condition (mean RT = 762 ms, SD = 142). The other condition (the RBGY set of keys) showed an intermediate speed (mean RT = 809 ms, SD = 145). Fisher's PLSD post hoc tests confirmed the differences between each combination of conditions (critical difference = 31 ms; -43 ms $<$ mean differences $<$ –90 ms; $.02 > p > .0001$).

There was also a significant interactive effect between response type and condition, $F(4, 352) = 4.4$, $p = .002$. As shown in

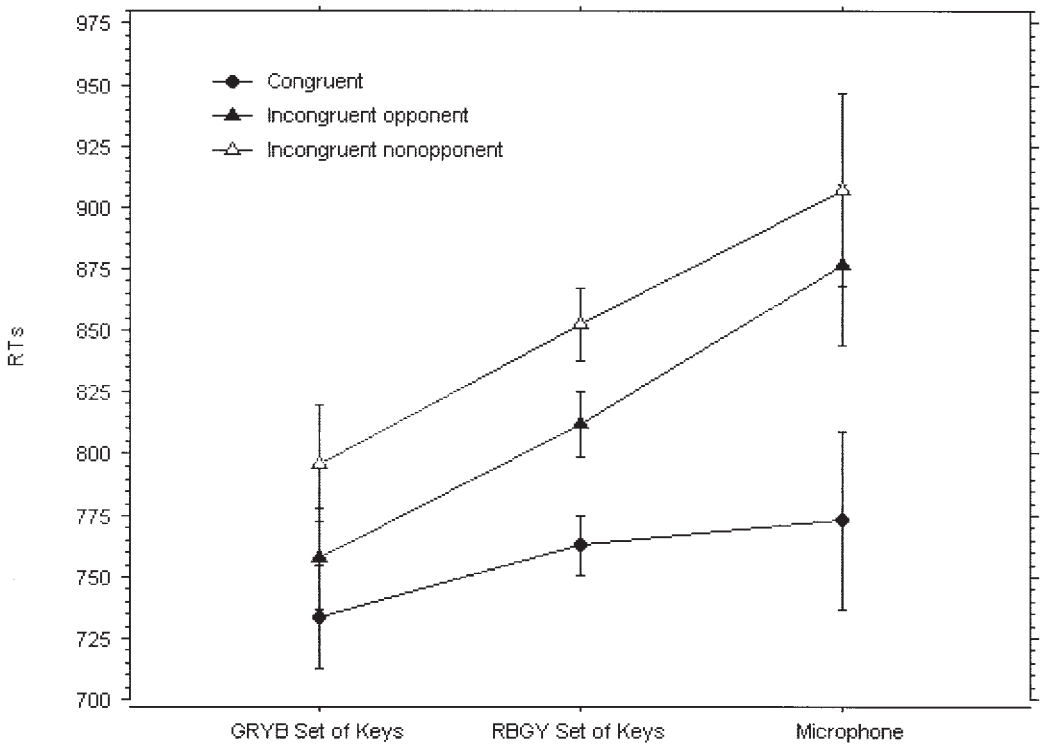

Figure 1. Experiment 1: Means (symbols) and standard errors (bars) of response times (RTs) in milliseconds to color–word pairs (congruent, incongruent opponent, incongruent nonopponent) when matching color patches labeled on computer keys in two different orderings (i.e., the green, red, yellow, blue [GRYB] and the red, blue, green, yellow [RBGY] sets) and by naming aloud the color (i.e., microphone). Standard errors were based on the across-participants variance of the scores for each condition separately.

Figure 1, the differences in RTs between conditions were greater when participants responded with a microphone than when they responded with key presses. However, the difference between opponent and nonopponent color–word pairs did not interact with response type. To further explore these effects, we performed two separate ANOVAs. The first analysis specifically explored the effects of opponent and nonopponent color–word pairs and response type on RTs, thus confirming that these two variables did not have a significant interactive effect on RTs, $F(2, 179) = 0.14$, $p = .87$. The second ANOVA explored the effect of response type on the strength of the Stroop effect (i.e., the difference in RTs between the averaged incongruent conditions and the congruent condition). This ANOVA showed that microphone responses led to the strongest interference (mean Stroop effect = 103 ms). Intermediate interference was found with arbitrary mapping of keys to colors (i.e., for the RBGY set of keys, mean Stroop effect = 71 ms), and the weakest Stroop effect was found with the more consistent mapping of keys to colors, where adjacent keys mapped onto opponent colors (i.e., for the GRYB set of keys, mean Stroop effect = 48 ms), $F(2, 179) = 8.6$, $p = .0003$. Fisher's PLSD post hoc tests confirmed that all conditions differed significantly from one another (19 ms $<$ critical difference $<$ 26 ms; -23 ms $<$ mean differences $<$ –54 ms; $.02 > p = .0001$).

To evaluate if the individual color–word pairs had different impacts on the degree of Stroop interference, we performed an additional repeated-measures ANOVA with each word–color pair (i.e., the four congruent pairs, the four opponent pairs, and the eight nonopponent pairs) as separate levels of one within-subject variable and mean RTs as the dependent variable. This ANOVA showed that the word–color pair was significant, $F(15, 2055) = 6.2$, $p = .0001$. Fisher's PLSD post hoc tests confirmed that all four congruent word–color pairs differed significantly from all of the opponent and nonopponent pairs (critical difference = 41 ms; 41 ms $<$ mean differences $<$ 118 ms; $.0001 < p = .05$). In addition, the post hoc tests failed to reveal any significant difference among the congruent word–color pairs (4 ms $<$ mean differences $<$ 32 ms). The post hoc tests also confirmed that several of the opponent word–color pairs differed significantly from nonopponent word–color pairs (i.e., *BLUE*–yellow and *GREEN*–red versus *RED*–blue and *RED*–yellow; *YELLOW*–blue versus *BLUE*–red, *GREEN*–yellow, *RED*–blue, and *RED*–yellow; *RED*–green versus *RED*–blue and *RED*–yellow; critical difference = 41 ms; 41 ms $<$ mean differences $<$ 60 ms; $.04 < p < .05$). Finally, the post hoc tests failed to reveal significant differences among the opponent word–color pairs (1 ms $<$ mean differences $<$ 8 ms) or among the nonopponent word–color pairs (1 ms $<$ mean differences $<$ 37 ms). Figure 2 illustrates the mean RTs for each individual word–color pair.

Individual differences in color vision. In the FMH test, the mean error rate was 57.8 ($SD = 29.4$). Performance on the test

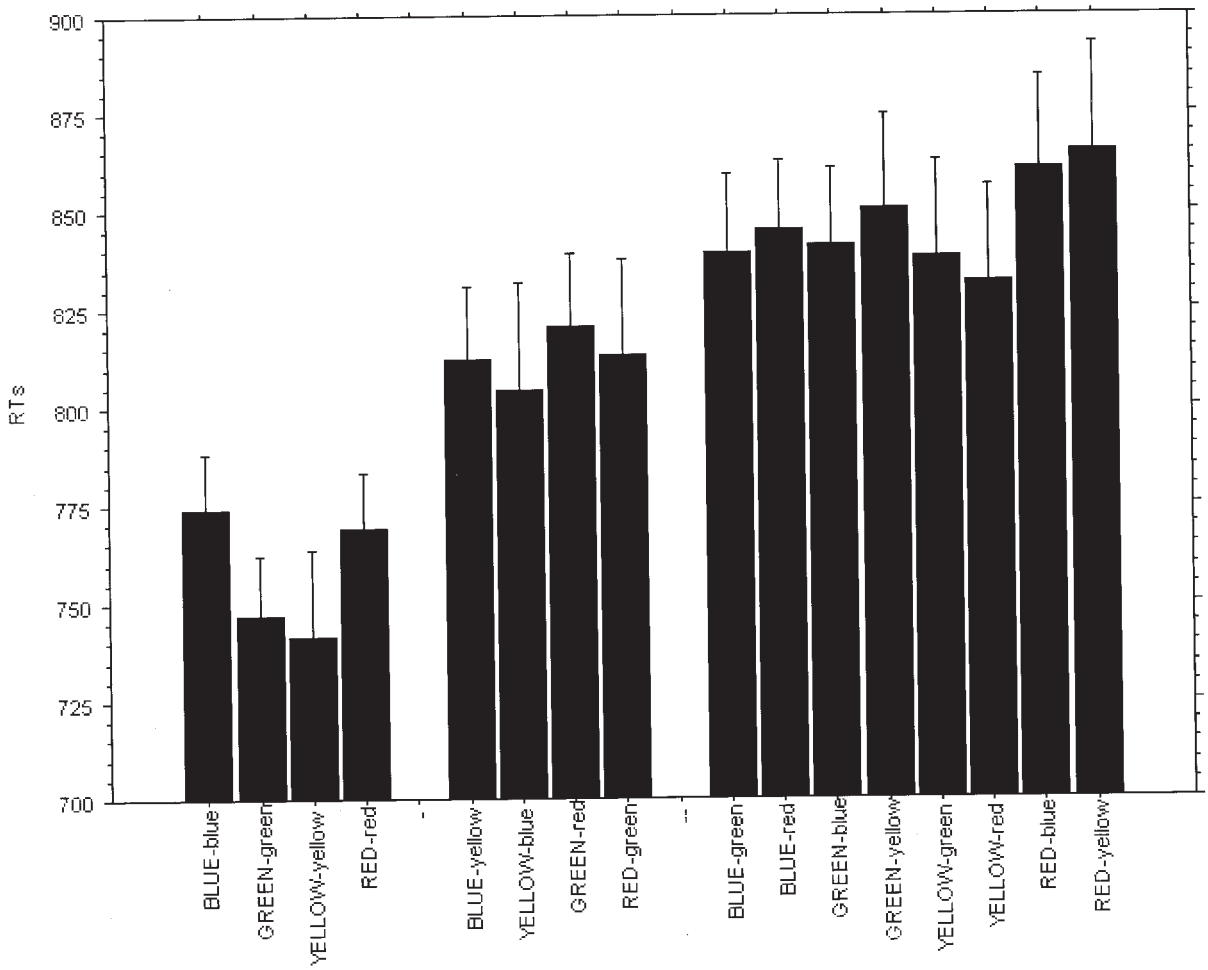

Figure 2. Experiment 1: Means (columns) and standard errors (bars) of response times (RTs) in milliseconds to all pairs of color–word combinations. Standard errors were based on the across-participants variance of the scores for each condition separately.

ranged widely (from 12 to 156). Four participants, all men, scored more than 2.5 standard deviations higher than the whole group's mean. An ANOVA with sex as the fixed variable and FMH scores as the random variable revealed that female participants had lower error rates (mean error score = 49, SD = 22) than the male participants (mean error score = 69, SD = 34), $F(1, 138) = 18.9$, $p = .0001$. For the Ishihara plates, the mean error rate was 1.03 (SD = 1.2), and scores ranged from 0 to 7. Four participants, 3 men and 1 woman, scored over 4. An ANOVA with sex as the fixed variable and scores on the Ishihara test as the random variable almost reached significance, $F(1, 138) = 2.9$, $p = .08$. Again, female participants had lower error rates (mean error score = 0.9, SD = 1.0) than the male participants (mean error score = 1.2, SD = 1.4). Because the aim of the study was to evaluate individual differences, all participants (including outliers) were included in the following analyses.

A simple regression between error scores for the Ishihara plates (as the regressor) and error scores in the FMH (as the dependent variable), revealed a significant positive relation and a correlation of moderate strength between the two color vision tests ($R = .39$, $p = .0001$).

Individual differences in color vision and the Stroop effect. First we performed a simple regression between error scores in the FMH test as the regressor and Stroop strength as the dependent variable. This revealed a significant relation (see Figure 3) and a correlation of moderate strength between color vision abilities and the Stroop effect ($R = .41$, $p = .0001$). Separate regression analyses between error scores in the FMH and Stroop strength were also performed on the female and male participants. These were performed to assess whether the previously observed correlation could have been mediated by the participants' sex (given the observed sex differences in color sensitivity and RTs). However, both sexes showed positive correlations (for women, $R = .30$, $p = .008$; for men, $R = .44$, $p = .0004$).

To evaluate whether the speed of color naming was influenced by individual differences in color perception, we also performed separate simple regressions with error scores in the FMH test as the regressor and mean RTs for congruent, opponent, or nonop-

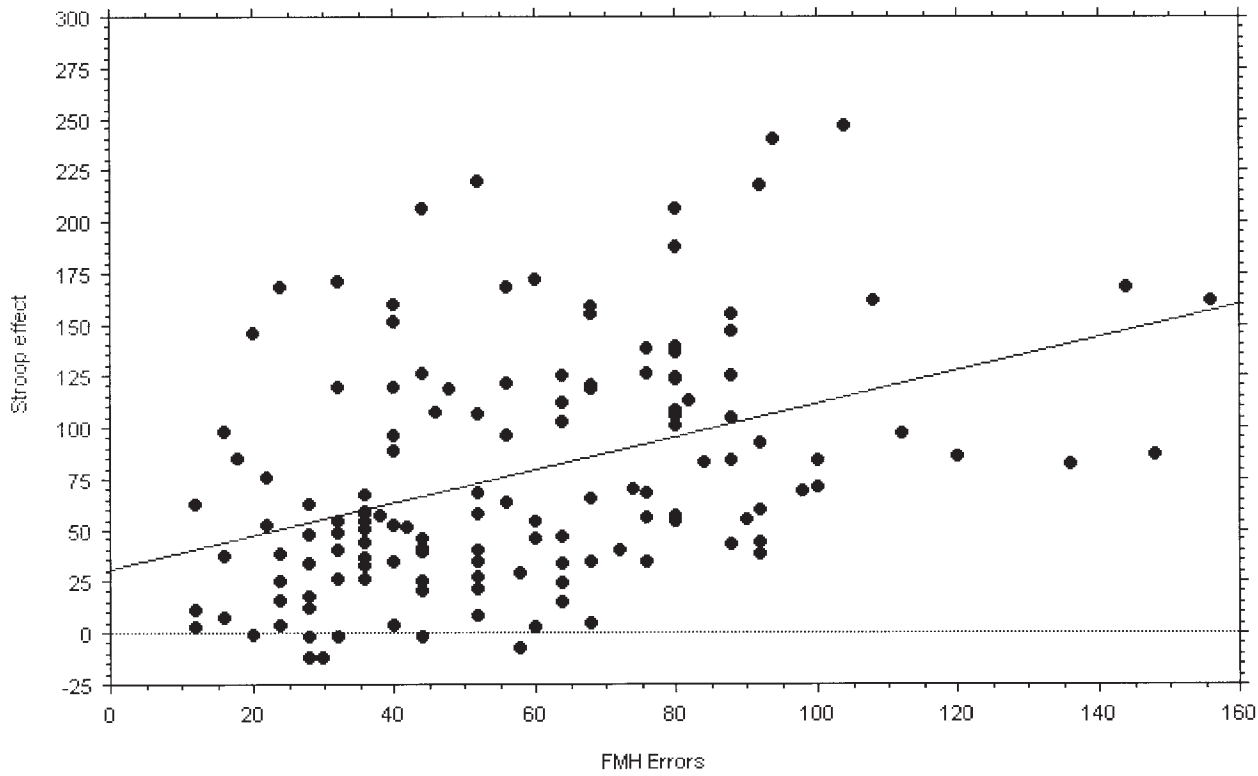

Figure 3. Experiment 1: Regression plot of individual differences in color vision, as measured by the Farnsworth-Munsell 100-Hue (FMH) Test error scores and the Stroop effect (i.e., the difference between all the color–word incongruent trials and the congruent trials). $Y = 31.04 + .804 \times X$; $R^2 = .17$.

ponent color–word pairs as the dependent variables. The regression on congruent trials revealed a significant relation and a correlation of moderate strength ($R = .37$, $p = .0001$). The regression on opponent trials revealed a significant relation and a correlation of moderate strength ($R = .38$, $p = .0001$). The regression on nonopponent trials revealed a significant relation and a correlation of moderate strength ($R = .42$, $p = .0001$). When the same regression analyses were performed between error scores in the Ishihara plates and strength of the Stroop effect on RTs to either opponent or nonopponent trials, these failed to reveal any significant relationship ($.05 < R < .08$).

Finally, our account would also predict that individuals showing specific difficulties in color discriminations within a particular region of color space (e.g., for colors within the blue quadrant of the FMH color circle) would also show increased Stroop interference for those trials in which the problematic color was either visually presented or referred to by the word (e.g., *BLUE*–yellow, *YELLOW*–blue, *BLUE*–red, *RED*–blue, etc.). To test this hypothesis, we first identified those participants who had a FMH error score greater than 2 standard deviations from the mean of all participants for colors within each quadrant (i.e., each of the four sets of color caps) of the FMH test. We then computed, for each of these individuals, the mean error rates (in percentages) and the Stroop strength (incongruent mean RTs minus congruent mean RTs) for all trials that included that color, either as the word's ink or as the word's name (or both for congruent trials). Table 1 shows the color-specific mean error rates and strength of the Stroop effect on RTs (bottom and top frames, respectively) of each group of participants with poor color vision within a particular color quadrant: Stroop interference in both percentage of errors and RTs was in all cases greater than the mean of all participants; most important, the largest increase in Stroop interference (about twofold for RTs) occurred for those conditions that included the color for which a group of participants had shown abnormal perception. Moreover, for these participants, Stroop interference was greater for the trials that included their deficient color as the word's ink than for the trials that included the color as the word. A repeated-measures ANOVA with color deficiency (blue, green, red, yellow) as the between-subjects variable, color stimulus (ink, word) as the within-subject variable, and strength of the Stroop effect on RTs (incongruent mean RTs minus congruent mean RTs) as the dependent variable confirmed the presence of a significant interactive effect of color deficiency and color stimulus, $F(9, 105) = 55.1$, $p = .0001$. A similar repeated-measures ANOVA with mean error rate (in percentages) as the dependent variable showed the same significant interactive effect of color deficiency and color stimulus, $F(9, 105) = 12.2$, $p = .0001$.

Experiment 2

In the previous experiment, there were equal numbers of congruent and incongruent trials ($N = 48$); that is, each congruent trial

Table 1
Experiment 1: Averaged Stroop Effect for Correct Trials in Which Red, Yellow, Green, or Blue Appeared as One of the Four Colors or Color Terms

	Red		Yellow		Green		Blue	
Participants	Word	Ink	Word	Ink	Word	Ink	Word	Ink
				RTs (ms)				
Red deficient ($N = 9$)	109	126	77	91	69	73	80	92
Yellow deficient ($N = 7$)	62	75	121	132	72	60	81	89
Green deficient ($N = 11$)	79	93	80	81	112	155	67	90
Blue deficient ($N = 12$)	68	75	71	75	80	95	115	168
All participants	83	69	60	68	67	83	78	58
				Mean errors (%)				
Red deficient ($N = 9$)	3.7	5.1	0.5	0.1	0.1	0.2	0.5	0.3
Yellow deficient ($N = 7$)	0.7	0.1	4.6	6.3	0.9	0.8	0.9	0.3
Green deficient ($N = 11$)	0.9	0.2	0.1	0.6	4.1	6.0	0.3	1.4
Blue deficient ($N = 12$)	0.6	0.4	0.3	0.6	1.4	2.1	6.6	8.1
All participants	0.3	0.7	0.1	0.5	0.4	0.8	1.1	1.2

Note. Participants included had error scores on the Farnsworth-Munsell 100-Hue Test (FMH) that were greater than two standard deviations from the mean scores of all participants for each color quadrant (red, yellow, green, or blue) of the FMH color circle.

(e.g., *BLUE* in blue) was repeated 12 times. Moreover, there were more repetitions of individual color–word pairs in the opponent condition (for which each ink color has only one word with which it can be paired; e.g., *BLUE* in yellow) than in the nonopponent condition (where *BLUE* can be paired with both red or green); in this way, both incongruent conditions would have the same number of trials ($N = 24$). This way of pairing dimensions of colors and words to form a stimulus set is typical of most Stroop research and, as Melara and Algom (2003, p. 442) pointed out, its popularity might stem from the investigators' intention to balance the number of items in the congruent and incongruent conditions so that the equal frequency of stimuli in the two conditions will not be predictive of the color or vice versa. However, Melara and Algom have shown that this assumption is incorrect because the relatively low dimensional uncertainty of each congruent stimulus encourages participants to expect congruent matchings, which enhances congruent performance relative to incongruent performance. Jacoby, Lindsay, and Hessels (2003) have also shown that the influence of word reading on Stroop color naming decreases as a function of the proportion of test items that are incongruent.

The above reasoning can also be extended to the present study's unequal repetitions of individual color–word pairs in the opponent and nonopponent conditions. In fact, the suspicion arises that the reduced interference of the opponent trials compared with the nonopponent condition might result from the relatively lower uncertainty of each opponent incongruent stimulus relative to a nonopponent incongruent stimulus. Hence, to rule out the possibility that the opponent advantage was simply due to the design that we had borrowed from the typical Stroop study, in Experiment 2 we again performed a single-trial computerized version of the classic color–word Stroop task, but the words paired to each color for the opponent and nonopponent incongruent matches were now equally predictive. In other words, each individual pair in the opponent condition (e.g., *BLUE* in yellow or *RED* in green) had the same number of repetitions ($N = 9$) as each individual pair (e.g., *BLUE* in red or *RED* in yellow) in the nonopponent condition. Moreover, each congruent color–word pair (e.g., *RED* in red) was also repeated for the same number of trials ($N = 9$).

If the reduced Stroop interference for opponent color–words matches observed in Experiment 1 was simply due the unequal frequency of specific color–words combinations, then in the following experiment we should observe no difference in the size of the Stroop interference between the opponent and the nonopponent conditions. In addition, we should also observe a general decrease of Stroop interference, because of the larger number of incongruent than congruent trials (cf. Jacoby et al., 2003).

Method

Participants. Thirty-five students (mean age = 22; $SD = 4$) at the University of Tromsø (24 women and 11 men) volunteered to participate in the new color–word Stroop experiment. None of these participants received the color vision tests.

Stimuli. These were the same as those used in the previous experiment.

Apparatus and procedure. As in the previous experiment, the *G, H, J,* and *K* keys were relabeled with focal blue, green, red, and yellow color patches. Only one sequence of the color patches was used in this experiment (from left to right): green, red, yellow, and blue (i.e., the GRYB set). The procedure was identical to the one for the previous experiment. However, the total number of trials was 144, and the frequency of each congruent color–word pair, incongruent opponent color–word pair, and incongruent nonopponent color–word pair was the same ($N = 9$). The trials' presentation order was completely randomized by use of SuperLab software.

Results

We calculated descriptive statistics for each participant, obtaining mean RTs for correct responses and mean error rates (in

percentages) for each combination of the variables word color (i.e., blue, green, red, and yellow), word name (i.e., *BLUE, GREEN, RED,* and *YELLOW*). RTs from trials on which errors occurred were excluded from analyses of the RTs (congruent: mean error = 3.8%, *SD* = 4.5; opponent: mean error = 4.4%, *SD* = 4.9; nonopponent: mean error = 3.3%, *SD* = 3.1), and trials with RTs greater than 3 standard deviations from each individual's mean RT for that cell were treated as outliers and excluded from all subsequent analyses. The number of trials excluded as outliers did not exceed 1% in any of the three conditions.

A repeated-measures ANOVA with condition (congruent, opponent, nonopponent) as the within-subject variable, sex as the between-subjects variable, and RT as the dependent variable revealed a main effect of condition, $F(2, 66) = 13.1$, $p = .0001$. Participants evaluated the congruent color–word pairs (mean RT = 732 ms, *SD* = 139) faster than opponent pairs (mean RT = 761 ms, *SD* = 107), whereas nonopponent pairs were evaluated the slowest (mean RT = 790 ms, *SD* = 123; see Figure 4 for an illustration of these results). There was no significant main effect, and there were no interactions of sex as a variable. Because the differences between the opponent and nonopponent conditions (or the congruent and the other two conditions) were predicted from the previous experiment's results, we tested the differences directly with separate ANOVAs. One analysis showed that the congruent and opponent conditions differed significantly, $F(1, 34) = 4.3$, $p = .05$. An additional ANOVA confirmed that the opponent condition was significantly faster than the nonopponent condition, $F(1, 34) = 12.9$, $p = .001$. Three additional ANOVAs, each comparing individual color–word pairs within the three main Stroop conditions (congruent, opponent, and nonopponent), failed to reveal significant differences. Finally, a repeated-measures ANOVA on mean percentage of errors as the dependent variable, condition (congruent, opponent, nonopponent) as the within-subject variable, and sex as the between-subjects variable showed no significant effects; for example, for condition, $F(2, 66) = 1.8$, $p = .18$.

Discussion

Although each individual color–word match appeared equally often in Experiment 2, this experiment replicated the reduced Stroop interference for opponent color–word pairs that we had observed in Experiment 1. However, there was a reduction in the size of the Stroop effect between Experiment 1 (in the condition with the RGBY sequence of color-labeled keys; mean RT difference = 71 ms) and Experiment 2 (mean RT difference = 41 ms). Such a decrease in the influence of word reading on Stroop color selection was likely due to the asymmetry in the proportion of incongruent items in Experiment 2, which had more incongruent (opponent and nonopponent combined) than congruent trials, versus the equal number of trials in the two conditions of Experiment 1 (cf. Jacoby et al., 2003; Lindsay & Jacoby, 1994; Melara & Algom, 2003; Tzelgov, Henik, & Berger, 1992). In addition, there was also a reduction of 10 ms in the size of the difference between the nonopponent and opponent conditions in Experiment 1 (mean RT difference = 38.3 ms, *SD* = 89) and Experiment 2 (mean RT

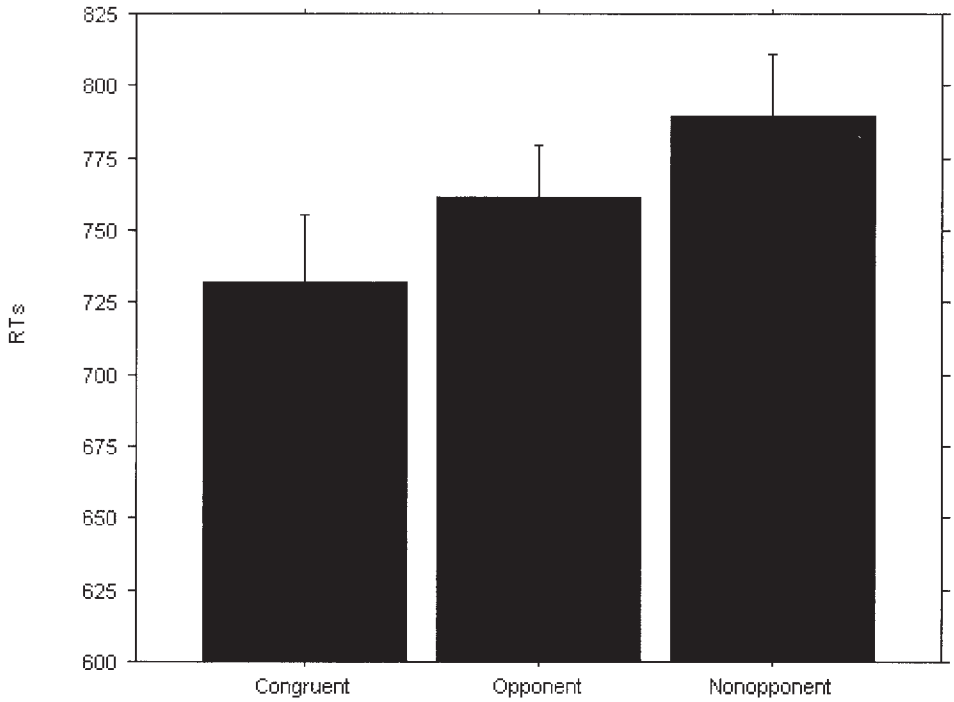

Figure 4. Experiment 2: Means (columns) and standard errors (bars) of response times (RTs) in milliseconds in each condition (congruent, opponent, and nonopponent). Standard errors were based on the across-participants variance of the scores for each condition separately.

difference = 28.6 ms, *SD* = 89). In summary, the two experiments have shown that the distance in similarity between the color dimension of the stimulus and the color dimension indicated by the word is a robust variable in Stroop interference and that this cannot be entirely accounted for by differences in the conditional probability of words and their matching colors. In fact, Klopfer (1996) had already shown that color naming becomes increasingly difficult when colors and words refer to closer-by regions of color space, and in Klopfer's study the color words were all equally predictive of the colors of the stimuli (as were the stimuli of Experiment 2 of the present study).

Finally, one negative finding from Experiment 1 was particularly suggestive: Although the two manual response mappings and the vocal response sets differed in the overall magnitude of the Stroop effect, there was no statistically significant difference in Stroop effect between opponent and nonopponent color–word pairs across the different types of response sets. The absence of such an effect of response sets can be seen as additional evidence that the difference in Stroop between the opponent and nonopponent color–word pairs is indeed primarily a perceptually based phenomenon. That is, if the sensory coding effects were having an influence at the level of response selection, an interaction of the type found for the overall Stroop effect would be expected for the opponent–nonopponent difference as well. However, there was no indication of such an interactive effect and, therefore, no support for an account favoring output instead of input factors on the opponent–nonopponent difference.

Neural Networks Simulations

In this experiment, we made a few simple modifications to a well-known parallel distributed processing model of the classic Stroop color–word task (Cohen et al., 1990) to (grossly) implement human trichromacy. We expected that such a manipulation would be sufficient to reproduce the reduced Stroop strength for opponent color–name pairs relative to nonopponent pairs. The basic architecture of the original model by Cohen et al. (1990) is displayed in Figure 5.

Cohen et al.'s (1990) model has two processing pathways, one for processing color information, the other for processing word information. Both pathways converge on a common response mechanism. In each pathway, there is an input layer, an intermediate layer, and an output layer. Also, there are two task demand units, which are used to model the allocation of attention to one or the other task (color naming or word reading). Processing occurs in the following way: Units at the input level are activated. Activation then feeds forward through the intermediate units and to the output units. A response occurs when the activation value of a response unit exceeds a certain threshold. Individual stimuli and responses have discrete, local representations in this model; that is, a single input unit in the color pathway, the word-reading pathway, and response pathway represents each ink color, color word, and response, respectively.

To simulate the experimental findings described above with opponent versus nonopponent color–word pairs, we introduced some modifications to Cohen et al.'s (1990) model. First, we needed to represent more than just two colors. Second, to implement, albeit at a gross level, three types of retinal cones, which are maximally sensitive to different wavelengths of light (i.e., the routine trichromacy of most primate species; cf. Surridge, Osorio, & Mundy, 2003), we introduced distributed representations for the ink-color inputs (where each unit in the input layer contributes to the representation of all colors) instead of the discrete, local representations (where one unit corresponds to one color or color word) used by Cohen et al. In other words, by coding ink colors as patterns of activation across different types of cones at the input layer, the network would encode similarities among colors in a biologically plausible fashion. For word inputs and outputs, we retained the local representations. We call this model the *human eye model.* The representations used in the current simulations are listed in Table 2, and the architecture of our slightly revised model is displayed in Figure 6.

In this scheme, each of the three color input units of the network corresponds to one of the three types of cones: the leftmost to the short wavelength cones, the middle to the medium wavelength cones, and the rightmost to the long wavelength cones. The listed representations were chosen on the basis of the fact that red mostly engages the long wavelength cones, green mostly engages the medium wavelength cones, blue mostly engages the short wavelength cones, and yellow engages medium and long wavelength cones to a roughly equal degree.

Training the Network

Training was done in the same manner as with the original model, with a few minor changes. A backpropagation learning algorithm (Rumelhart, Hinton, & McClelland, 1986) was used to produce the correct response to information in one of the two processing pathways. Thus, training patterns consisted of input to one of the two task-demand units, as well as input to the corresponding pathway. Note that the training stimuli do not contain any instances of simultaneous ink-color and color–word inputs.

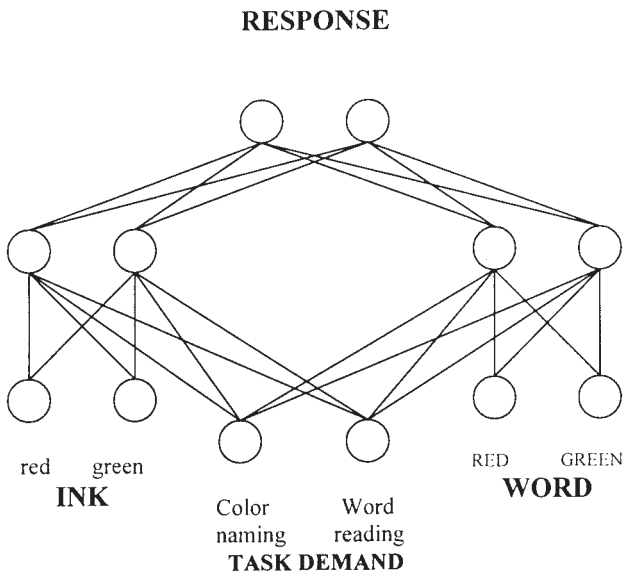

Figure 5. The architecture of the original connectionist model of Stroop performance of Cohen et al. (1990).

Table 2
Representations Used in the Trichromatic Stroop Model

Input	Vector
Color	
Red ink	0 0 1
Green ink	0 1 0
Blue ink	1 0 0
Yellow ink	0 1 1
Word	
Red	1 0 0 0
Green	0 1 0 0
Blue	0 0 1 0
Yellow	0 0 0 1
Response	
Red	1 0 0 0
Green	0 1 0 0
Blue	0 0 1 0
Yellow	0 0 0 1

This is meant to reflect the fact that people rarely see Stroop-like stimuli, except in experimental tasks.

When training started, the connection strengths of the network were all set at small random values. All connections' strengths were modifiable by backpropagation, except for those between the task demand units and the intermediate units in their respective pathways. These strengths were frozen at 4.0, to offset a permanent bias of −4.0 applied to all intermediate units. Thus, when a training pattern was presented to one of the pathways, the input at the task demand unit effectively brought the activation of the intermediate units in the corresponding pathway up to 0, which is in the middle of their most dynamic range. This arrangement is similar to that in Cohen et al.'s (1990) model and was meant to capture the effective filtering of an attentional mechanism. Similarly, we accepted their assumption that reading is a more automated process than color naming; thus, 83.3% of all training patterns were color words, whereas only 16.7% of all training patterns were ink-color inputs. Each training sweep consisted of five instances of each color word and one instance of each ink color. We used a learning rate of 0.1. Each instance of the model was trained for 60,000 sweeps, at which point the output units' activation very rarely deviated from the correct value by more than 0.2.

Another difference between the current model and that of Cohen et al. (1990) was that, in the original model, the activation value of a given unit was not based on its instantaneous net input but, rather, on the net input averaged over time. In the current model, output activations are computed in a single step, and responses are evaluated on the basis of the root mean square error of the output patterns. That is, a given output pattern of the model is compared with the desired output. The root of the mean of the squared deviations of such a comparison provides a measure of the model's performance. Hence, like other authors (e.g., Seidenberg & McClelland, 1989), we interpreted the size of the root mean square error as our estimate of the latency of the model's response. Indeed, output activations computed in this way can be shown to correspond to the asymptotic activations in a cascaded network.

Simulation 1: The Basic Stroop Effects

The purpose of our model's first simulation was to see whether it was able to replicate the basic Stroop phenomena modeled by Cohen et al. (1990). Specifically, (a) that word reading is faster than color naming, (b) that color naming is influenced by words, (c) that word reading is not affected by ink color, and (d) that there is less facilitation than interference. Fifty instances of the model were trained, and the model was tested for word reading alone, color naming alone, and Stroop color naming. The root mean square errors for each output pattern were then calculated and averaged across model instances. The model behaved in accordance with the basic original findings, displaying an advantage for word naming as well as the Stroop interference with incongruent pairs and the facilitation effects with congruent ones.

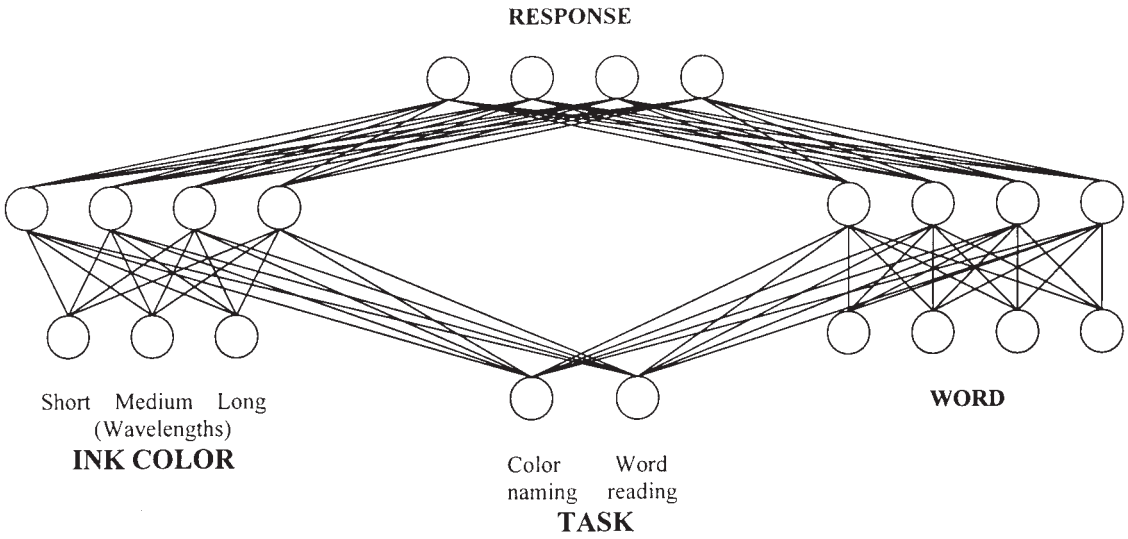

Figure 6. The architecture of the current trichromatic model.

Simulation 2: Human Eye Model: Reduced Stroop Strength With Opponent Color–Word Pairs

The main goal was to simulate the attenuated Stroop interference, observed with human participants, when the color–word pairs were opponent colors as opposed to nonopponent colors. As expected, the interference effect was smaller when the incongruent colors were opponent colors, compared with when they were nonopponent colors. Each instance of the model ($N = 50$) was treated as a subject in an ANOVA with condition (congruent, opponent, nonopponent colors) as the within-subject variable and mean root mean square error as the dependent variable. Note that each instance of the model before training had a different random configuration of connection strengths. The analysis revealed a significant difference, $F(2, 49) = 442.0$, $p = .0001$. Post hoc Fisher's PLSD comparisons confirmed that each condition differed significantly from any other (critical difference = .015; .06 < mean differences < .19; $p = .0001$).

Simulation 3: Alien Eye Model

To evaluate whether the attenuation of the interference effect obtained in the previous simulation was really due to the manipulation at the input layer, and not just to a side effect of some other variation of Cohen et al.'s (1990) original model, we implemented an expanded version using a localist representation scheme for the ink-color inputs (i.e., one unit representing each color). The ink-color representation vectors used in the simulation were 0001, 0010, 0100, and 1000 for red, green, blue, and yellow, respectively.

The network was then trained as before. Again, the root mean square errors were calculated and averaged across the 50 model instances for each condition. The model displayed the basic Stroop phenomena: facilitation for congruent colors and interference for noncongruent colors. However, this model did not yield the reduced Stroop effect with opponent colors. Although the same ANOVA, with instances of the model used as subjects, condition (congruent, opponent, nonopponent colors) as the within-subject variable, and mean root mean square error rates as the dependent variable, revealed a significant effect, $F(2, 49) = 381.0$, $p = .0001$, the post hoc Fisher's PLSD comparisons showed that the effect was entirely due to the congruent condition being significantly different from both incongruent conditions (critical difference = .014, mean differences = .14 and .15, $p = .0001$), with no difference between the opponent and nonopponent conditions (mean difference = .008). Figure 7 illustrates the results of both Simulation 2 (human eye) and Simulation 3 (alien eye).

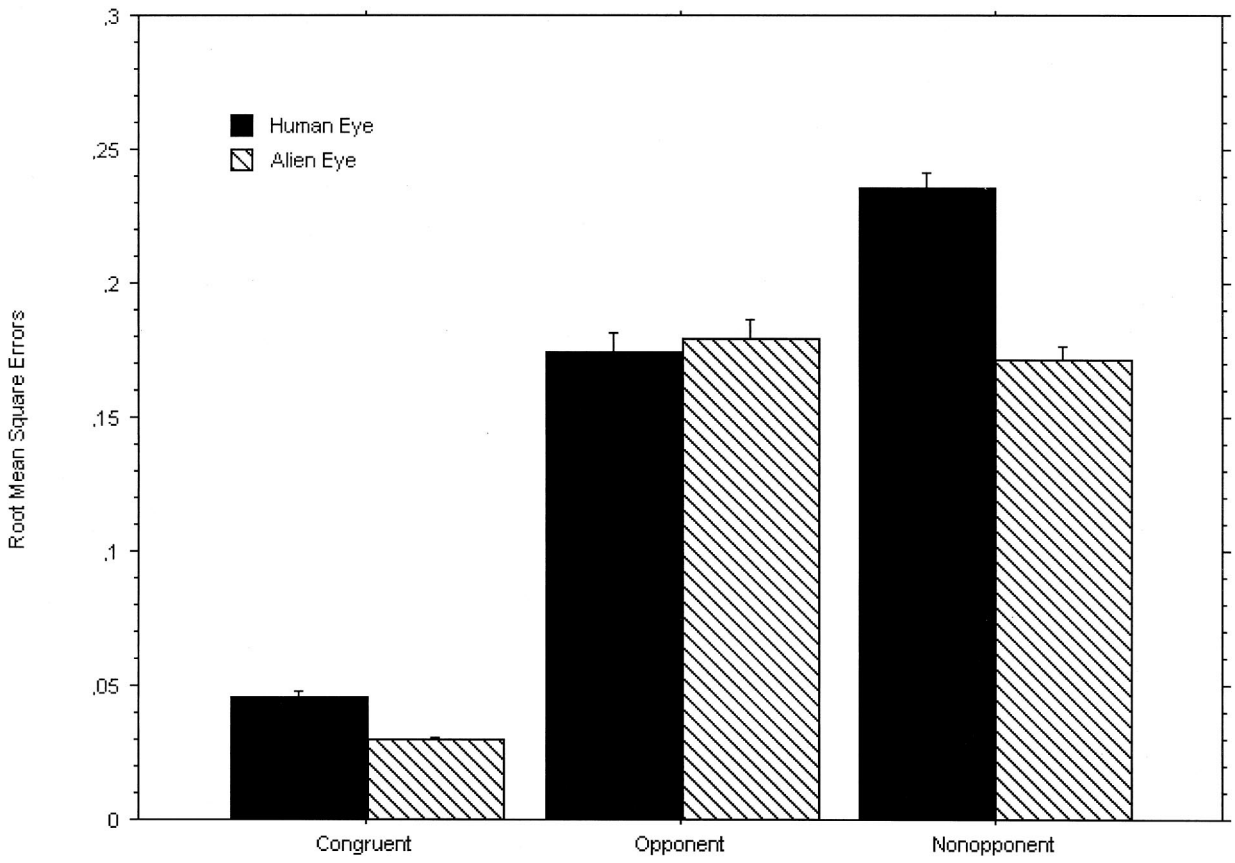

Figure 7. Root mean square errors (95% confidence intervals as bars) of Simulation 2, the alien eye model (hatched columns), and Simulation 3, the human eye model (filled columns).

We conclude that the effect of opponent pairs evident in the previous simulation derived from the similarity structure imposed by the specific organization of the input vectors. In the model, the major difference is due to the color yellow, and this change is apparently sufficient to produce a similarity structure in the human eye model that differs dramatically from that of the alien eye model. That is, the geometrical distance between the four input vectors is modified between the two models. The patterns of activation across the intermediate units for the different ink-color inputs indicate that the model has captured this new similarity structure. In the human eye model, the intermediate units develop representational patterns reminiscent of color opponency, with each unit being either on for red and off for green or off for red and on for green, as well as on for blue and off for yellow or off for blue and on for yellow. In the control (alien) model, the pattern shown by the intermediate unit is less clear-cut, and what tends to happen is that a given color activates only one of the intermediate units.

General Discussion

We found that (a) individual participants' differences in their color discrimination abilities influenced their performance in the classic Stroop color–word task; (b) this was also reflected in the effect of sex of the participants on both color vision and Stroop interference; (c) color–word pairs with opponent colors reduced the strength of Stroop interference (i.e., the difference in RTs with their corresponding congruent trials) compared with nonopponent color–name pairs; (d) the reduced strength of Stroop interference with opponent colors is robust, and the effect remains strong when the conditional probabilities of words matching specific colors are made equal; and (e) vocal and manual responses led to different Stroop interference: The strongest interference was found with verbal responses and the weakest was found when participants matched keypresses to color patches with opponent colors adjacent to one another.

The association between individual differences in color vision and Stroop performance is of particular interest and underlines a fundamental factor of the Stroop effect. Indeed, the present results point to the need to incorporate sensory input factors into a complete account of the Stroop effect. One way to conceptualize the influence of sensory input factors is to posit that individual differences in sensory discriminations change the landscape of color space. That is, assuming that the color (e.g., green) and the word (naming a different color, e.g., *BLUE*) activate different regions in color space, these regions might differ among individuals, both in extent and degree of overlap. Hence, the word *blue* would interfere to a lesser degree in individuals who are able to discriminate well between blue and green than in individuals who make poor discriminations between these two colors. This hypothesis was supported by the positive relationship between susceptibility to Stroop interference and the error rates in a color-discrimination task. In addition, individuals with poor discriminations of adjacent colors (e.g., blue and green) were more susceptible to Stroop interference when the color or the word was within the deficient range of the color spectrum than were participants who were not deficient at discriminations of the same range of colors.

These findings suggest that interference can be determined at an early selection stage, that is, the perceptual-encoding stage of the ink-color information. This aspect is perhaps made more explicit by the neural network simulations. In such simulations, we introduced a (coarsely) biologically plausible modification of a well-developed model of the Stroop effect (i.e., Cohen et al., 1990; cf. O'Reilly & Munakata, 2000), so that the encoding vector or input layer coded ink colors as patterns of activation across different types of "cones." First, such a simple change at the level of the input layer was sufficient to simulate the reduced Stroop interference observed with human participants. The human eye or trichromatic model showed a significant difference between incongruent color–word pairs involving opponent colors as opposed to incongruent color–word pairs involving nonopponent colors. Only this distributed trichromatic input layer appeared responsible for generating the difference in effect magnitude for opponent and nonopponent color–word pairs. The lack of a difference in interference between types of incongruent color–word pairs for the alien eye or hyperchromatic model strengthens the interpretation that organization at the encoding stage is sufficient for the emergence of similarity effects on Stroop performance. Note that, according to the model, the color and the word pathways in the model do not meet until the response level of the model; therefore, the differences in effect magnitudes between opponent and nonopponent pairs in these simulations can be attributed entirely to the processing of physical color. Thus, the difference between the two levels of interference appears to reflect the similarity structure imposed at the color-encoding stage (i.e., the number of scalars shared) of the trichromatic model.

We surmise that individual differences in sensory discriminations (e.g., genetically based variations) are likely to shape different color spaces. In multidimensional scaling studies, individuals with color vision deficiency yield, in a color-matching task, a "degenerate" version of Newton's color circle with the deficient sides of the circle collapsed together (Shepard, 1997; Shepard & Cooper, 1992). In other words, for different individuals, the color and the word might activate regions of color space of differing extent and degree of overlap. In turn, the use of color labels will be influenced by such differences among individuals of colors' proximity within color space. For example, a person with difficulty in discriminating different shades in one region of the spectrum would be expected to be more prone to (wrongly) use the name of an adjacent hue (cf. Lindsey & Brown, 2002). This would also lead to increased lexical competition (cf. MacWhinney, 1987), that is, increased applicability of terms in the lexicon other than the target one. In the color–word Stroop, an increase in perceptual difficulty would result in increased interference when such an alternative, potentially applicable, name is provided by the word stimulus. Thus, one proposal is that early perceptual processing might produce its effects by cascading onto processes that influence access to color semantics and the production of color labels. As perceptual encoding begins, information flows onto the subsequent levels, affecting later processing. Increased difficulty of encoding will have effects transmitted up the system from one level to the next, which reveal themselves as a strong source of interference.

Nevertheless, one major source of interference in color–word Stroop has been taken to occur at a conceptual level (Klein, 1964; Klopfer, 1996; Luo, 1999; Scheibe, Shaver, & Carrier, 1967;

Stirling, 1979). From our perspective, we would surmise that input factors have an organizational effect on conceptual/semantic space; consequently, perceptual encoding (sensory input from the color system) could be a necessary component of any complete account of the Stroop effect. Indeed, several studies have shown that the difficulty in naming colors increases directly with the semantic proximity of the words in which the colors were printed. Klein (1964) was the first to show that distantly related color names produced a decrease in Stroop interference compared with closely related color names. In addition, common words that did not name colors caused a further decrease in interference, but words that named objects implying specific colors (e.g., *lemon, sky*) caused a smaller decrease in interference than did words without such associations (e.g., *friend*) or rare words and nonsense syllables. Klein also proposed that semantic relatedness modulates the attention-catching power of the word; the greater the competition between the color concept activated by the word and the color itself, the more the participant needs to perceptually resample the color (in Klein's, 1964, words, the participant "restimulates himself with the color of the word–color combination," p. 585). Thus, in Klein's view, it would be the seeking of additional perceptual input from the color that would then result in a lengthening of RTs. More recently, Klopfer (1996) has shown that Stroop interference is not the same for all incongruent color–word pairs (e.g., the interference obtained with *PURPLE* when the color of the word is yellow is less than the interference with *PURPLE* in blue). Klopfer pictures this process as both the color and the word (naming a different color) activating two regions within a color conceptual space. Distance in color space and degree of overlap between such regions will determine the extent of processing (e.g., selective attention) needed to discriminate between the two. Near neighbors in color space are then associated with increased performance time.

One of the other findings of our empirical study was that female participants evaluated color–word pairs faster than male participants and showed less Stroop interference. Given that the female participants had overall better color vision than the male participants (a common finding in large samples of participants; e.g., Pickford, 1951), it seems consistent that the women were also faster than the men. However, in his thorough review of the literature on Stroop effect, MacLeod (1991) concluded, "There are no sex differences in Stroop interference at any age" (p. 184). Given the present results, we tend to agree with the beginning of MacLeod's next sentence in the article: "Perhaps this is too strong" (p. 184).

Modulatory effects of response types on Stroop performance have been described in various studies (e.g., Henik, Ro, Merrill, Rafal, & Safadi, 1999; Logan & Zbrodoff, 1998; White, 1969). MacLeod (1991) concluded that the Stroop effect is greater, if not exclusively present, with vocal responses (cf. Besner, 2001; Durgin, 2000; McClain, 1983). Yet Stroop effects with manual responding could reflect the mediation of subvocal color naming (and might disappear with sufficient trials, as suggested by Dalrymple-Alford & Azkoul, 1972). Nevertheless, one of the interesting results of the present study was that different manual responses led to different Stroop interference: The strongest was with the RBGY mapping and the weakest was with the GRYB mapping (i.e., when matching keypresses to color patches with opponent colors adjacent to one another). These effects are generally consistent with structural correspondence effects of the type reported by Miller (1982) and Proctor and Reeve (1985). Note that in the GRYB mapping, adjacent keys map onto points of the color circle that are an average of 150° apart (green-to-red = 180° apart, red-to-yellow = 90°, yellow-to-blue = 180°), whereas with the RBGY mapping, the difference between adjacent keys is 90°. Hence, this effect may reflect the influence of conceptual proximity on Stroop interference. Remarkably, because it seems unlikely that the typical participant is using an explicit knowledge that green and red or yellow and blue are on opposite sides of the color circle, such an influence might be occurring at a rather implicit level of processing.

Other modulatory effects on Stroop performance that should be taken into consideration are those due to similarities and differences between the phonetic aspects of the two words, as well as differences in length of the string of letters or syllables. In our Norwegian version of the Stroop task, *RED* is *RØD*, *YELLOW* is *GUL*, *BLUE* is *BLÅ,* and *GREEN* is *GRØNN.* Hence, all words have a different syllable, though *GUL* and *GRØNN* share the initial phoneme. However, our findings showed no differences between color–word pairs within each condition (congruent, opponent, and nonopponent), which would suggest that the trials including *GUL* and *GRØNN* as words were not treated differently than the other pairs. Similarly, regarding word lengths, the differences are minimal in Norwegian (where all words except one have three letters) compared with English (where all the words have different lengths). Among the nonsignificant findings, it was surprising that the error rates with the Ishihara plates did not correlate significantly with Stroop interference because (a) FMH error rates and Ishihara error rates were correlated and (b) a study by van Boxtel, ten Tusscher, Metsemakers, Willems, and Jolles (2001) using the (clinical) cards version of the Stroop test had shown that red–green color weakness led to slower color naming. Nevertheless, there was a tendency for a positive correlation with Ishihara errors and size of the Stroop effect. To conclude, the FMH test may have better predictive power than the Ishihara test for RT measurements on other color-related tasks. Future studies could assess in detail the effects of specific color impairments on Stroop interference. For instance, participants with color-vision deficiency who are missing the M cone would have problems discriminating the red from the green opponent color, as well as discriminating these from the intermediate nonopponent colors in the yellow–orange range. We would then expect to observe in such participants not only an increased Stroop interference compared with nondeficient participants but also a reduced difference between incongruent nonopponent Stroop interference and incongruent opponent Stroop interference. Future simulations may implement even more biologically plausible constraints than the simple one we used here. For example, color opponency—that is, $L - M$ and $S - (L + M)$ mechanisms—may have cascade effects that are somewhat different, and their study could further refine the predictive power of the models. Thus, if red and green inhibit each other early in the visual system (i.e., early at the level of the retinal ganglion cells, as well as later at the cortical level; De Valois & De Valois, 1997; Engel, 1999), the influence of such inhibitory effects should be measurable in the production of a color label in a Stroop task.

References

Abboud, H., & Sugar, D. (1990). SuperLab (Version 1.04) [Computer software]. Phoenix, AZ: Cedrus.

Besner, D. (2001). The myth of ballistic processing: Evidence from Stroop's paradigm. *Psychonomic Bulletin & Review, 8,* 324–330.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition "automatic"? *Journal of Experimental Psychology: General, 131,* 220–240.

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.

Dalrymple-Alford, E. C., & Azkoul, J. (1972). The locus of interference in the Stroop and related tasks. *Perception & Psychophysics, 11,* 385–388.

De Valois, R. L., & De Valois, K. K. (1997). Neural coding of color. In A. Byrne & D. R. Hilbert (Eds.), *Readings on color: Vol. 2. The science of color* (pp. 93–140). Cambridge, MA: MIT Press.

Durgin, F. H. (2000). The reverse Stroop effect. *Psychonomic Bulletin & Review, 7,* 121–125.

Engel, S. A. (1999). Using neuroimaging to measure mental representations: Finding color-opponent neurons in visual cortex. *Current Directions in Psychological Science, 8,* 23–27.

Farnsworth, D. (1957). *The Farnsworth-Munsell 100-Hue Test for the examination of color vision.* Baltimore: Munsell Color Company.

Henik, A., Ro, T., Merrill, D., Rafal, R., & Safadi, Z. (1999). Interactions between color and word processing in a flanker task. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 198–209.

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology, 83,* 299–303.

Ishihara, S. (1951). *Tests for color blindness.* London: Lewis.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review, 10,* 638–644.

Jensen, A. R., & Rohwer, W. D., Jr. (1966). The Stroop color–word test: A review. *Acta Psychologica, 25,* 36–93.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132,* 47–70.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General, 129,* 340–360.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77,* 576–588.

Klopfer, D. S. (1996). Stroop interference and color–word similarity. *Psychological Science, 7,* 150–157.

Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1998). The effects of irrelevant stimuli: The time course of stimulus–stimulus and stimulus–response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 688–714.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219–234.

Lindsey, D. T., & Brown, A. M. (2002). Color naming and the phototoxic effects of sunlight on the eye. *Psychological Science, 13,* 506–512.

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 978–992.

Lu, C.-H., & Proctor, R. W. (2001). Influence of irrelevant information on human performance: Effects of S-R association strength and relative timing. *Quarterly Journal of Experimental Psychology Human Experimental Psychology, 54*(A), 95–136.

Luo, C. R. (1999). Semantic competition as the basis of Stroop interference: Evidence from color–word matching tasks. *Psychological Science, 10,* 35–40.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.

Mantere, K., Parkkinen, J., Mäntyjärvi, M., & Jasskelainen, T. (1995). Eigenvector interpretation of the Farnsworth-Munsell 100-Hue Test. *Journal of the Optical Society of America, 12,* 2237–2243.

McClain, L. (1983). Effects of response type and set size on Stroop color–word performance. *Perceptual and Motor Skills, 56,* 735–743.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110,* 422–471.

Miller, J. O. (1982). Discrete versus continuous stage models of human information processing: In search of partial output. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 273–296.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

Phaf, R. H., Van der Hejden, H. C., & Hudson, P. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

Pickford, R. W. (1951). *Individual differences in colour vision.* London: Routledge & Kegan Paul.

Proctor, R. W., & Reeve, T. G. (1985). Compatibility effects in the assignment of symbolic stimuli to discrete finger responses. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 623–639.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110,* 88–125.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PSP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45–76). Cambridge, MA: MIT Press.

Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology, 15,* 462–471.

Scheibe, K. E., Shaver, P. R., & Carrier, S. C. (1967). Color association values and response interference on variants of the Stroop test. *Acta Psychologica, 26,* 286–295.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Shepard, R. N. (1997). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In A. Byrne & D. R. Hilbert (Eds.), *Readings on color: Vol. 2. The science of color* (pp. 311–356). Cambridge, MA: MIT Press.

Shepard, R. N., & Cooper, L. A. (1992). Representations of colors in the blind, color-blind, and normally sighted. *Psychological Science, 3,* 97–104.

Sichel, J. L., & Chandler, K. A. (1969). The color–word interference test: The effects of varied color–word combinations upon verbal response latency. *Journal of Psychology, 72,* 219–231.

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 28,* 643–661.

Surridge, A. K., Osorio, D., & Mundy, N. I. (2003). Evolution and selection of trichromatic vision in primates. *TRENDS in Ecology and Evolution, 18,* 198–205.

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition, 20,* 727–735.

van Boxtel, M. P. J., ten Tusscher, M. P. M., Metsemakers, J. F. M., Willems, B., & Jolles, J. (2001). Visual determinants of reduced performance on the Stroop color–word test in normal aging individuals. *Journal of Clinical and Experimental Neuropsychology, 23,* 620–627.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13,* 304–319.

White, B. W. (1969). Interference in identifying attributes and attribute names. *Perception & Psychophysics, 6,* 166–168.

Zhang, H., & Kornblum, S. (1998). The effects of stimulus–response mapping and irrelevant stimulus–response and stimulus–stimulus overlap in four-choice Stroop tasks with single-carrier stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 3–19.

Zhang, H., Zhang, J., & Kornblum, S. (1998). A parallel distributed processing model of stimulus–stimulus and stimulus–response compatibility. *Cognitive Psychology, 38,* 386–432.

Received July 8, 2003
Revision received October 7, 2004
Accepted October 29, 2004 ■

ORDER FORM

Start my 2005 subscription to ***JEP: Human Perception & Performance*!** ISSN: 0096-1523

___ $140.00, **APA Member/Affiliate** _____
___ $287.00, **Individual Nonmember** _____
___ $698.00, **Institution** _____
In DC add 5.75% / In MD add 5% sales tax _____
Total Amount Enclosed $ _____

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.

AMERICAN PSYCHOLOGICAL ASSOCIATION

SEND THIS ORDER FORM TO:
American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Or call **(800) 374-2721**, fax **(202) 336-5568**.
TDD/TTY **(202) 336-6123**.
For subscription information, e-mail:
subscriptions@apa.org

◯ **Send me a FREE Sample Issue**
◯ **Check enclosed** (make payable to APA)
Charge my: ◯ VISA ◯ MasterCard ◯ American Express

Cardholder Name _____
Card No. _____ Exp. Date _____

_____
Signature (Required for Charge)

BILLING ADDRESS:
Street _____
City _____ State _____ Zip _____
Daytime Phone _____
E-mail _____

SHIP TO:
Name _____
Address _____
_____
City _____ State _____ Zip _____
APA Member # _____ *XHPA15*