

# The effect of Stroop interference on the categorical perception of color

ALISON J. WIGGETT

Bangor University, Bangor, Wales

AND

IAN R. L. DAVIES

University of Surrey, Guildford, England

In two experiments, we examined the effects of Stroop interference on the categorical perception (CP; better cross-category than within-category discrimination) of color. Using a successive two-alternative forced choice recognition paradigm (deciding which of two stimuli was identical to a previously presented target), which combined to-be-remembered colors with congruent and incongruent Stroop words, we found that congruent color words facilitated CP, whereas incongruent color words reduced CP. However, this was the case only when Stroop interference was presented together with the target color, but not when Stroop stimuli were introduced at the test stage. This suggests that target name, but not test name generation, affects CP. Target name generation may be important for CP because it acts as a category prime, which, in turn, facilitates cross-category discrimination.

The perceptual system processes a vast amount of visual information at any given time. We encounter objects, see other people, witness events, and, somehow, are able to make sense of it all. To achieve this, the incoming information has to be put into manageable chunks, or categories. The process of segmenting the world into categories means that nonidentical stimuli can be treated as equivalent, rather than each instance having to be treated as new and unique, a process that, according to Rosch (1978), is done for the purpose of *cognitive economy*. The case of color is a particularly good example of this. Although we can discriminate millions of colors, we divide the color spectrum into a small number of categories (~10), and a few basic color terms are sufficient to describe all possible colors (Thompson, 1995). Furthermore, there is evidence to suggest that colors (and many other perceptual categories) are not only labeled, but also *perceived* categorically.

## Categorical Color Perception

Categorical perception (CP) is defined by better discrimination across category boundaries than for equivalently separated stimuli within the same category (Harnad, 1987). The earliest demonstrations of CP were in speech perception. Given a continuum of sounds varying from one phoneme to another in equal increments (e.g., /ba/ to /da/), sounds straddling a category boundary are more discriminable than sounds within a category (Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1973, 1991). Subsequent research showed evidence for CP also

in visual perception. The advantage of cross-category over within-category discrimination has been reported in the perception of many visual categories—for example, facial expressions (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Etcoff & Magee, 1992; Young et al., 1997), familiar faces (Beale & Keil, 1995), and familiar objects (Newell & Bühlhoff, 2002). In color perception, CP has been reported across a range of tasks, including recognition memory (Uchikawa & Shinoda, 1996), *same/different* judgments (Bornstein & Korda, 1984; Boynton, Fargo, Olson, & Smallman, 1989; Özgen & Davies, 2002), and similarity judgments (Laws, Davies, & Andrews, 1995).

Although the generality and robustness of the category effects just described are generally accepted, there are competing accounts of the origin of the effects. Naturalistic theories (e.g., Bornstein, 1987; Snowden, 1987) propose that CP is the result of inherent perceptual processes that produce discontinuities in perceptual representations. That is, the mapping of a uniform physical continuum onto its perceptual representation produces local discrimination maxima around which category boundaries form. According to naturalistic theories, CP is both inborn and perceptually mediated. Perceptual change theories (e.g., Harnad, Hanson, & Lubin, 1991) propose that learning to categorize stimuli (through practice or through learning to distinguish stimuli by name) results in a change to the stimulus representation. This could lead to perceptual change in the sensitivity around boundaries (e.g., Harnad et al., 1991) or to new perceptual structures (Schyns & Rodet, 1997). Perceptual change theories are similar to

naturalistic theories in assuming that CP effects are perceptually mediated but differ in postulating that the effects are learned, rather than inherent. Direct language theories (Kay & Kempton, 1984; Roberson & Davidoff, 2000) propose that comparison of some form of linguistic representation, such as a label, contributes to discrimination decisions. Since stimuli that straddle a category boundary have different labels but stimuli from the same category share a label, this additional source of information would confer an advantage on cross-category decisions.

Roberson and Davidoff (2000) tested this direct language account using a delayed color discrimination task, in which a target color (e.g., blue1) was followed after an interstimulus interval (ISI) by two test stimuli, one of which was identical to the target and one of which (the foil) was different. The foil could be either from the same category as the target (e.g., blue2) or from an adjacent category (e.g., green1). Crucially, the perceptual distance between the target and the foil was the same in both conditions. During the ISI, there was no interference, verbal interference (reading a list of words), or visual interference (tracking a line on screen), and the task was to decide which of the test stimuli was physically identical to the target. They found a significant CP effect in the no-interference condition, and this survived visual, but not verbal, interference. This suggests that verbal interference disrupted the comparison of linguistic codes, eliminating this advantage to cross-category decisions, resulting in all the decisions' having to rely on the comparison of visual codes. Comparison of linguistic codes does not have to be overt to produce a cross-category advantage. It could be *implicit* (Gilbert, Regier, Kay, & Ivry, 2006; Kay & Kempton, 1984), perhaps involving some prephonological representation, such as the *lemma* (Caramazza, 1997). Our use of terms or phrases, such as labeling or the online use of language, is intended to include either explicit or implicit linguistic influences.

Consistent with the implicit influence of language, category effects also occur in simultaneous discrimination tasks in which there is no requirement to remember a target color across the ISI, as in a delayed discrimination task. For instance, using a visual search task, Gilbert et al. (2006) found shorter search times when target and distractors were from different categories than when they were from the same category, but only for targets presented in the right visual field (RVF). This RVF category effect was eliminated by a concurrent verbal interference task. They interpreted their results as showing that implicit online use of language influences target detection (see also Kay & Kempton, 1984; Winawer et al., 2007). Furthermore, Pilling, Wiggett, Özgen, and Davies (2003) suggested that although labeling may be involved in CP, its function could be to activate a category code that supports the decision process in delayed color discrimination tasks (Bornstein & Korda, 1984; Pilling et al., 2003). This idea is closely related to earlier work on priming by Rosch (1975) and Neumann and D'Agostino (1981). These studies suggested that priming with a color word in color discrimination tasks results in a mental representation that is visual

in nature but not as specific as the actual physical color code. The code could be activated either *top down* through labeling or *bottom up* by the color (the category exemplar). This dual activation by the target and the target label could produce a more durable, or more precise, memory code than either alone, thus facilitating comparison with the test colors. As with comparison of category labels, deciding that two colors are different would be easier if they evoked different category codes than if they evoked the same category code, resulting in color CP. Now, however, just labeling the target may be sufficient to produce CP, particularly with long ISIs, because of the enhanced robustness of the memory code. In the experiments presented here, one of our approaches to exploring the nature of color CP was to test this *priming account* of CP, using Stroop-like stimuli in color discrimination tasks.

Our second approach was to test the effect of reducing the memory load by shortening the interval between target and test (ISI) from 5,000 msec, as in Roberson and Davidoff (2000) and Pilling et al. (2003), to 1,000 msec. Since a visual code is considerably attenuated by 5 sec (e.g., Pilling, 2001; Posner & Keele, 1967), reliance on language may be greater than with shorter ISIs. If language is one source of CP, reducing the ISI should reduce CP. However, providing the target label, as in a congruent Stroop stimulus (e.g., GREEN paired with a green stimulus), should reinforce category code activation and facilitate discrimination for categorically different stimuli.

### **Stroop Interference**

To test the role of language in CP, we investigated the effects of disrupting the generation of the linguistic code, using Stroop interference. In the classic Stroop color-word task, subjects have to name the color of the "ink" in which an incompatible color name is printed. Stroop (1935), and many researchers since (see MacLeod, 1991, for a review) have shown that this incongruity of stimulus dimensions leads to a decrease in the accuracy and speed with which the color can be named. Many variations of the original task have been found to be equally susceptible to Stroop interference. However, the vast majority of studies have focused on the relevant stimulus dimension (the word—i.e., the name of the ink color) and have not been interested in the other stimulus dimension (the color—i.e., the actual hue). It is likely that a color discrimination task with Stroop interference does not involve all the same mechanisms as the original Stroop task. Nevertheless, if color naming is involved in both tasks, both should be susceptible to Stroop interference.

Surprisingly little is known about how and whether Stroop interference affects or interacts with the visual representation of a color. Only a few studies have directly addressed these issues (Bradlyn & Rollins, 1980; Flowers, 1975; Flowers & Blair, 1976; Flowers & Dutch, 1976). Flowers and Blair, for example, conducted a series of experiments investigating the use of verbal and visual codes in color classification tasks with Stroop interference. The focus was on the hues of the to-be-remembered inks and the effect of varying the similarity of these hues. In a speeded

classification task, subjects had to sort six different colors into two groups of three colors. The colors that were to be grouped together were either adjacent in color space (e.g., red, orange, yellow vs. green, blue, purple) or nonadjacent (e.g., red, yellow, blue vs. orange, green, purple). The colors were presented either as XXXX patterns or as incongruent words (Stroop stimuli). Flowers and Blair found stronger Stroop interference when nonadjacent hues were grouped than when adjacent hues (e.g., red, orange, yellow) were grouped. They suggested that grouping nonadjacent hues requires access to verbal codes, whereas grouping adjacent hues is based on perceptual similarities (see also Flowers & Dutch, 1976). Perceptual decisions are not disrupted by verbal response competition—hence, no Stroop interference. We used a similar logic for the experiments presented here. Whether CP is susceptible to Stroop interference should depend on the type of coding underlying the advantage of cross-category over within-category discrimination: if CP is truly perceptual, Stroop interference should have no effect; if CP is due to the online influence of language, it should be susceptible to Stroop interference.

### The Present Experiments

To clarify the role of language in CP, we combined a delayed color discrimination task with Stroop interference. If CP requires naming and if Stroop weakens name generation, introducing Stroop interference at the various stages should expose the role of language in CP. In Experiment 1A, we used the same task as that in Roberson and Davidoff (2000) and Pilling et al. (2003), but by including congruent and incongruent Stroop interference with the target color, we explored the effects of inhibiting or facilitating target color naming. In Experiment 1B, the ISI and test stimuli presentation times were shortened, thus lowering memory loads. This should facilitate the use of visual codes and, if CP still occurred, would suggest that it had a perceptual component. In Experiments 2A and 2B, we explored the effects of inhibiting or facilitating naming of the test stimuli by introducing Stroop interference at the response stage. If CP is based solely on matching color labels, the two test stimuli, as well as the target, must be named correctly. If, however, target name generation facilitates CP by priming the category code, Stroop interference at test presentation should not affect CP.

## GENERAL METHOD

### Apparatus and Stimuli

The experiments were conducted on a 15-in. Sony Trinitron monitor controlled by a PC with a Diamond stealth graphics card. The chromaticity of all the color stimuli was checked using a Minolta CS-100 colorimeter. A high-resolution timer DLL produced by ExacTics was used to ensure accurate event timing. The subjects were screened for color vision problems, using the City University color vision test (Fletcher, 1980).

Although we used the perceptually uniform metric CIE  $L^*u^*v^*$  to equate separations between adjacent pairs of stimuli, we also will give their equivalent Munsell notations for those more familiar with that system (see Hunt, 1987, for more on CIE  $L^*u^*v^*$  and Munsell). Sixteen stimuli were used (eight blue and eight green) that combined four hues and four levels of lightness, all at Munsell chroma

7.4 (see Figure 1A). All horizontally adjacent stimuli in the matrix had a perceptual distance ( $\Delta E$ ) of nine units in CIE  $L^*u^*v^*$ ; the vertically adjacent stimuli differed by four units. Only the horizontally adjacent pairs were used in the discrimination task, giving four cross-category pairs and eight within-category pairs. Forced choice naming of the stimuli (*blue* or *green*) by 20 subjects showed the following percentages for naming: 1.44B *blue*, 97%; 8.76BG *blue*, 86%; 6.06BG *green*, 82%; 3.46BG *green*, 98% (see Pilling et al., 2003); lightness did not affect naming. Each stimulus was  $55 \times 55$  mm, subtending  $6.3^\circ$  of visual arc at an approximate viewing distance of 500 mm. The target stimulus was presented in the center of the monitor, whereas the two test stimuli were presented to the left and right of the monitor center, separated by 1 cm.

The label BLUE or GREEN or XXXX in black uppercase lettering (font size of 16) was shown centered on the target color (Experiments 1A and 1B) or test colors (Experiments 2A and 2B). The color words or XXXX measured  $15 \times 5$  mm ( $1.7^\circ \times 0.6^\circ$ ). All the stimuli of Munsell hue 1.44B and 8.76BG were classified as congruent when combined with the label BLUE and as incongruent when combined with the label GREEN. All the stimuli of 6.06BG and 3.46BG hue were congruent when labeled GREEN and incongruent when labeled BLUE.

### Procedure and Design

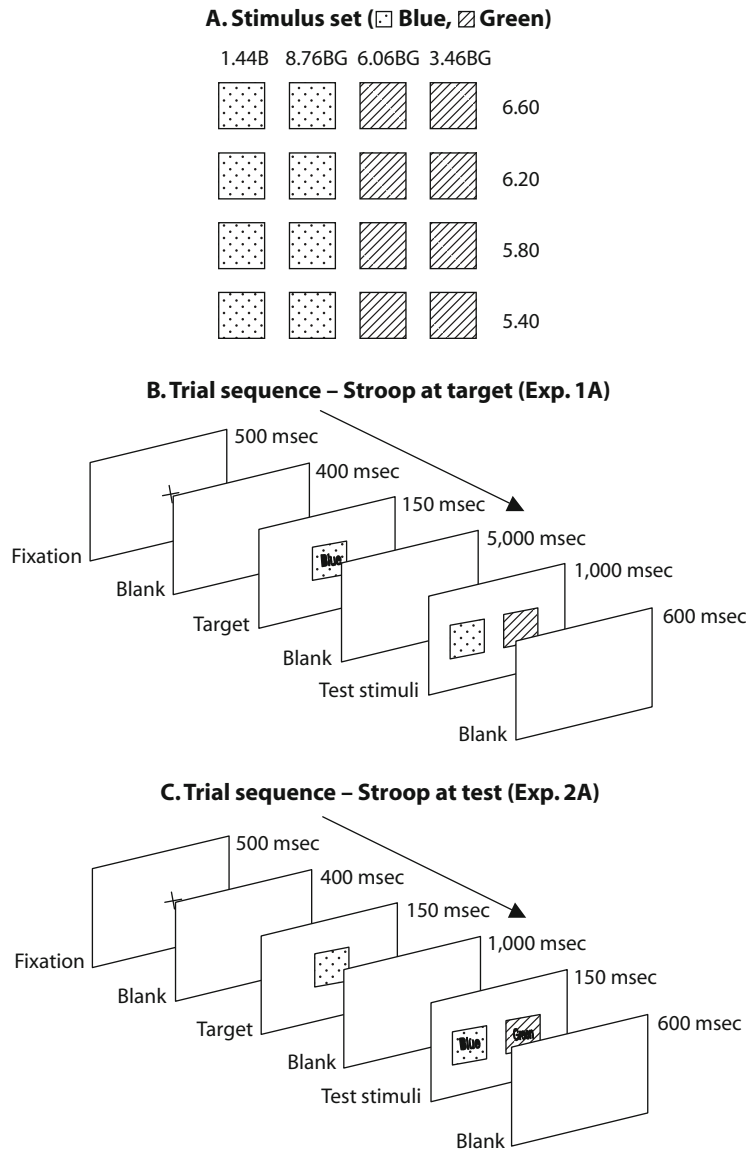
Discrimination was measured using a delayed two-alternative forced choice task. A target color was presented for a given time, followed by two test stimuli—the target plus a foil—after an unfilled interval. The task was to choose the color identical to the target. The test stimuli were either from the same category (within-category pairs) or from different categories (cross-category pairs). Stimulus pairs were chosen at random on each trial from the 12 horizontally adjacent pairs (Figure 1A), but all 12 pairs were shown with equal frequency in the three Stroop conditions. Each stimulus of the pair was shown as the target and the foil with equal frequency. At test presentation, the target was presented as the left or right stimulus with equal frequency. Responses were made by pressing the corresponding mouse button (left or right), and the instructions emphasized accuracy. Reaction times from the onset of the test stimuli until the response were recorded. The subjects were instructed to ignore the superimposed words and attend to the color. The intertrial interval was 600 msec. Three blocks of 10 practice trials were completed before starting the experimental trials, and auditory feedback was given for incorrect responses on these trials.

### Data Analysis

The data in all the experiments were analyzed as follows. The mean accuracy scores for within- and cross-category pairs in each Stroop condition were calculated. These were analyzed by a repeated measures two-way ANOVA with pair type (cross-category or within-category) and Stroop condition (neutral, congruent, or incongruent) as factors. Significant interactions were explored with post hoc analysis using protected *t* tests (two-tailed) and one-way ANOVAs on the size of the category effect (cross-within). We then repeated the analysis using median reaction times as a covariate (ANCOVAs). In all cases, the ANCOVAs showed the same pattern as the ANOVAs (see the Appendix), and we therefore will not report them in the Results sections.

## EXPERIMENTS 1A AND 1B

The main goal in our first experiment was to investigate the effects of labeling the target on CP by combining the target with a conflicting name, a congruent name, or neutral symbols. If CP requires target labeling, it should be weakened by the incongruent Stroop stimuli and, possibly, enhanced in the congruent condition. If congruent and incongruent Stroop interference do not selectively affect CP, it would suggest that language is not the origin of CP.



**Figure 1.** (A) Matrix of stimuli used. The stimuli were equally spaced in CIE (1976)  $L^*u^*v^*$ , varying on four hue and four value levels. There were four cross-category and eight within-category pairs. (B) Sequence of events of an experimental trial with Stroop interference at target presentation in Experiment 1A. In Experiment 1B, the interstimulus interval was reduced to 1,000 msec, and test stimuli presentation to 150 msec. (C) Sequence of events of an experimental trial used in Experiment 2A with Stroop at test presentation. In Experiment 2B, test stimuli presentation was increased to 300 msec.

## Method

**Subjects.** Twenty-eight subjects (6 of them male, 22 female; age range, 18–34 years) participated in Experiment 1A; a further 21 subjects (7 of them male, 14 female; age range, 20–38 years) completed Experiment 1B. All were students at the University of Surrey and participated in exchange for course credits.

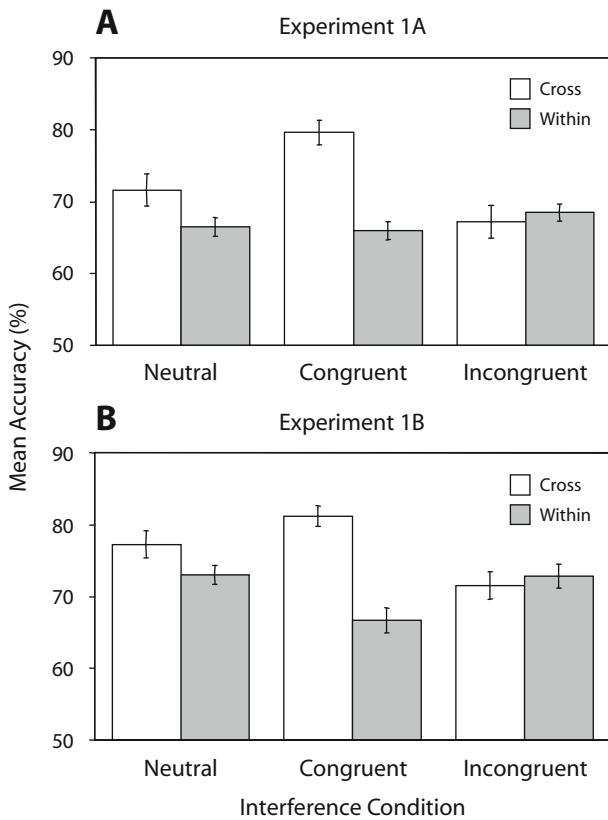
**Procedure and Design.** The trial sequence is shown in Figure 1B. At the start of each trial, a fixation cross was presented for 500 msec, followed 400 msec later by the target for 150 msec, followed by a 5,000-msec unfilled ISI, after which the two test stimuli (the target and the foil) were presented for 1,000 msec or until a response was made (Experiment 1A). There were 216 trials in total, presented

in six blocks of 36 trials. There were equal numbers of congruent, incongruent, and neutral trials in each block in random order.

In Experiment 1B, test stimuli presentation times and the length of the ISI were reduced, relative to Experiment 1A. The target (plus Stroop interference) was presented for 150 msec, followed by a 1,000-msec ISI, after which the two test stimuli were presented for 150 msec only.

## Results

The mean accuracy scores are presented in Figure 2A for Experiment 1A and in Figure 2B for Experiment 1B.



**Figure 2.** Mean discrimination accuracy (error bars show  $\pm 1$  standard error) for cross-category and within-category pairs for each interference condition (neutral, congruent, and incongruent) in Experiments 1A (A) and 1B (B).

Mean reaction times for all the experiments are presented in Table 1. In Experiment 1A, both main effects and the interaction were highly significant. Accuracy was higher for cross-category than for within-category pairs [ $F(1,27) = 19.72$ ,  $MS_e = 72.36$ ,  $p < .001$ ]. The significant interference effect [ $F(2,54) = 7.81$ ,  $MS_e = 46.50$ ,  $p < .01$ ] reflected lower mean accuracy in the neutral and the incongruent conditions

than in the congruent condition for cross-category comparisons [minimum  $t(25) = 4.41$ ,  $p < .001$ ]. The significant interaction [ $F(2,54) = 14.5$ ,  $MS_e = 54.43$ ,  $p < .001$ ] reflected the variation in cross-category advantage across the three Stroop conditions. There was an advantage of cross-category over within-category discrimination in the neutral [ $t(26) = 2.23$ ,  $p < .05$ ] and the congruent [ $t(26) = 6.02$ ,  $p < .001$ ] conditions, but not in the incongruent condition [ $t(26) = 0.55$ ,  $p > .05$ ]. This finding was further supported by a one-way ANOVA on the difference between cross- and within-category conditions (category effect), which showed a highly significant effect of Stroop condition [ $F(2,54) = 14.5$ ,  $MS_e = 108.8$ ,  $p < .001$ ]. The CP effect was significantly enhanced in the congruent condition, as compared with the neutral condition [ $t(25) = 3.09$ ,  $p < .01$ ].

In Experiment 1B, the ANOVA showed that accuracy was higher for the cross- than for the within-category pairs [ $F(1,20) = 18.18$ ,  $MS_e = 64.86$ ,  $p < .001$ ], and the interaction was also significant [ $F(2,40) = 12.12$ ,  $MS_e = 61.97$ ,  $p < .001$ ]. There was no interference effect [ $F(2,40) = 2.32$ ,  $MS_e = 41.98$ ,  $p > .05$ ]. There was an advantage of cross-category over within-category pairs in the neutral [ $t(19) = 1.8$ ,  $p < .05$ ] and the congruent [ $t(19) = 6.71$ ,  $p < .01$ ] conditions, but not in the incongruent condition [ $t(19) = 0.55$ ,  $p > .05$ ]. The one-way ANOVA on the difference between cross- and within-category conditions (category effect) showed a highly significant effect of Stroop condition [ $F(2,40) = 12.12$ ,  $MS_e = 123.93$ ,  $p < .001$ ]. As in Experiment 1A, the CP effect was significantly enhanced in the congruent condition, as compared with the neutral condition [ $t(18) = 3.14$ ,  $p < .01$ ]. A separate  $2 \times 3$  ANOVA across experiments, with experiment (1A or 1B) and Stroop condition (neutral, congruent, or incongruent) as factors, showed a significant effect of Stroop [ $F(2,94) = 26.48$ ,  $MS_e = 115.3$ ,  $p < .001$ ], but no effect of experiment and no Stroop  $\times$  experiment interaction (minimum  $p = .88$ ).

## Discussion

In Experiments 1A and 1B, a significant CP effect was found in the neutral conditions. Thus, despite the fact that

**Table 1**  
Mean Reaction Times (in Milliseconds, With Standard Errors) in Neutral, Congruent, and Incongruent Stroop Conditions for Cross-Category and Within-Category Discriminations in Experiments 1A, 1B, 2A, and 2B

Pair Type		Stroop Interference					
		Neutral		Congruent		Incongruent	
		<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1A	Cross category	871.10	33.50	849.10	28.20	874.50	32.20
	Within category	921.50	33.90	924.90	33.50	904.00	25.70
Experiment 1B	Cross category	764.98	29.20	756.19	23.80	764.67	26.30
	Within category	760.91	22.90	767.69	24.10	780.14	27.20
Experiment 2A	Cross category	738.88	23.50	743.13	25.80	737.15	25.10
	Within category	741.91	23.80	751.75	25.50	748.05	21.90
Experiment 2B	Cross category	768.18	25.93	728.23	25.40	762.26	24.80
	Within category	779.11	24.62	767.80	22.16	758.12	23.89

Note—The variation in reaction times reflects the difference in presentation times across experiments. In Experiment 1A, the test stimuli were presented for 1,000 msec; in Experiments 1B and 2A, for 150 msec; and in Experiment 2B, for 300 msec.

the perceptual distances between all the color pairs were held equal (in CIE  $\Delta E$ ), color discrimination was significantly better for cross-category pairs than for within-category pairs. This effect increased in the congruent conditions but disappeared in the incongruent conditions. This result suggests that being able to generate the correct target name facilitates cross-category discrimination and supports the idea that the advantage of cross-category discrimination may be based on comparison of category labels or priming of the category code.

If tasks that invite, or require, verbal coding are susceptible to Stroop interference, the results of Experiment 1B suggest that language is used even in the shortened version of the task. As in Experiment 1A, though, this pattern of results cannot distinguish between comparison of target or test labels and priming of the category code as influences on CP. To help distinguish between these two accounts, the next experiment used Stroop interference at test presentation.

## EXPERIMENTS 2A AND 2B

It seems likely that no CP effect was found in the incongruent condition in Experiments 1A and 1B because Stroop interference at target presentation weakened target name generation. This still leaves open the question of whether correct test stimuli labeling is also necessary for CP. If CP depends on the comparison of verbal codes or the online use of language in some form, the test stimuli must also evoke the appropriate linguistic code. However, if the function of the target label is to activate the category code, this might be sufficient to yield CP without labeling of the test stimuli.

Stroop interference was introduced at response by labeling the target and the foil either congruently or incongruently. CP could arise because the target (e.g., blue), as well as the test stimuli (e.g., one blue, the other green), are labeled, and the task is done by matching these labels in some form. If CP is based on such a comparison, congruent Stroop interference at the response stage should facilitate CP, but incongruent interference should reduce CP. However, if the target is labeled, reinforcing the category code, labeling of the test stimuli may not be necessary for CP, and it should still occur despite the interference.

### Method

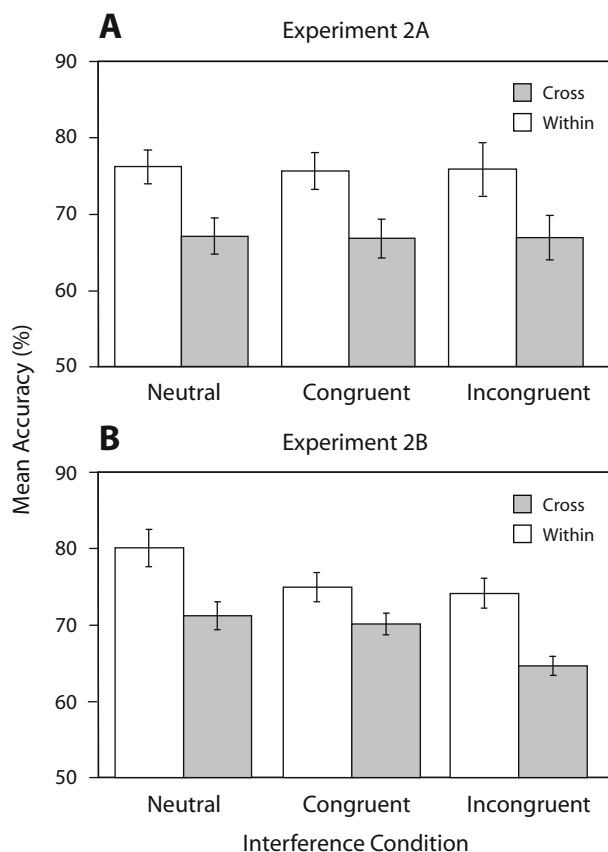
**Subjects.** Twenty-eight subjects (13 of them male, 15 female; age range, 19–46 years) participated in Experiment 2A; a further 24 subjects (6 of them male, 18 female; age range, 18–40 years) completed Experiment 2B. All were students at the University of Surrey and participated in exchange for course credits. None had participated in the previous experiments.

**Procedure.** The trial sequence is shown in Figure 1C. The trial sequence and stimulus presentation times were identical to those in Experiments 1A and 1B, except that Stroop interference was now presented concurrently with test stimuli. For incongruent trials, there were equal frequencies of both test stimuli being incongruent, the “correct” test stimuli (target) being incongruent, and the “incorrect” test stimuli (foil) being incongruent. The trials were split into six blocks of 40 trials, with interference type varying within a block. Test stimuli presentation time (plus Stroop interference) was increased in Experiment 2B (relative to Experiment 2A) to 300 msec.

### Results

Figure 3A shows the mean across-subjects accuracy scores for cross- and within-category trials, for each Stroop condition in Experiment 2A. It appears that the cross-category scores were higher than the within-category scores for all the conditions. This was supported by the ANOVA showing a significant main effect of pair type [ $F(1,27) = 18.64$ ,  $MS_e = 163.72$ ,  $p < .001$ ]. Neither the effect of interference [ $F(2,54) = 0.02$ ,  $MS_e = 52.23$ ,  $p > .05$ ] nor the pair type  $\times$  interference interaction [ $F(2,54) = 0.07$ ,  $MS_e = 57.8$ ,  $p > .05$ ] was significant. Post hoc tests showed that accuracy for cross-category pairs was significantly greater than that for within-category pairs in all the conditions [minimum  $t(26) = 1.88$ ,  $p < .05$ ]. The one-way ANOVA on the difference between cross- and within-category conditions (category effect) showed no significant effect of Stroop condition on the size of the CP effect [ $F(2,54) = 0.7$ ,  $MS_e = 115.58$ ,  $p > .05$ ].

Figure 3B shows the mean across-subjects accuracy scores for cross- and within-category trials for each Stroop condition in Experiment 2B. As with Experiment 2A, it appears that the cross-category scores were higher than the within-category scores for all the conditions. The ANOVA showed a significant main effect of pair type [ $F(1,23) = 29.72$ ,  $MS_e = 71.92$ ,  $p < .001$ ]. There was also a signifi-



**Figure 3.** Mean discrimination accuracy (error bars show  $\pm 1$  standard error) for cross-category and within-category pairs for each interference condition (neutral, congruent, and incongruent) in Experiments 2A (A) and 2B (B).

cant main effect of interference [ $F(2,46) = 9.84$ ,  $MS_e = 47.18$ ,  $p > .001$ ]; however, the pair type  $\times$  interference interaction [ $F(2,46) = 1.65$ ,  $MS_e = 46.69.8$ ,  $p > .05$ ] was not significant. Post hoc tests showed that accuracy for cross-category pairs was significantly greater than that for within-category pairs in all the conditions [minimum  $t(22) = 1.88$ ,  $p < .05$ ]. Furthermore, there was no significant effect of Stroop condition on the size of the CP effect [ $F(2,40) = 1.65$ ,  $MS_e = 93.37$ ,  $p > .05$ ] in the one-way ANOVA. The separate  $2 \times 3$  ANOVA across experiments, with experiment (2A or 2B) and Stroop condition (neutral, congruent, or incongruent) as factors, showed no significant main effects and no significant interaction (minimum  $p = .43$ ).

## Discussion

In Experiment 2A, cross-category discrimination was better than within-category discrimination in all the conditions, consistent with the perceptual code's surviving the short ISI and producing CP. Since there was no Stroop effect, this suggests that language is not involved in CP at the test stage. Experiment 2B was conducted to rule out the possibility that the test presentation times in Experiment 2A (150 msec) were too short to induce Stroop interference. The results clearly show that CP survived all interference conditions with 300-msec test presentation times.

## GENERAL DISCUSSION

The aim of the experiments reported here was to explore the nature of color CP in discrimination tasks. Our approach was twofold. First, we tested the matching-of-labels account of CP, using Stroop-like stimuli. Stroop interference was used because only tasks that require the use of or access to verbal labels should be susceptible to Stroop interference (Flowers & Dutch, 1976; MacLeod, 1991). Our second approach was to test the effect of reducing the memory load by shortening the interval between target and test from 5,000 msec (as used in Roberson & Davidoff, 2000) to 1,000 msec. The visual code should be considerably weakened by 5 sec, and so reliance on labeling may be greater with long ISIs than with shorter ISIs. We found that the advantage of cross-category over within-category discrimination accuracy was susceptible to Stroop interference presented concurrently with the target. Reducing the memory load did not reduce the effect of Stroop interference. However, Stroop interference concurrently with test presentation did not affect CP. This was the case for matched (relative to Experiment 1B) stimuli presentation times (Experiment 2A), as well as for matched Stroop exposure (Experiment 2B). These results (especially those of Experiments 1A and 1B), along with those of previous reports (Pilling et al., 2003; Roberson & Davidoff, 2000; Winawer et al., 2007), provide strong support against the idea that cross-category discrimination advantages (CP) in delayed color discrimination tasks are purely perceptual. The results of the experiments presented here support the suggestion made by Pilling et al. that target name generation may be necessary for CP in

these types of tasks. However, the results (especially those of Experiments 2A and 2B) also suggest that the direct language account of color CP is incomplete.

In Experiments 1A and 1B, we introduced congruent and incongruent Stroop interference with the to-be-remembered color. The results suggest that preventing naming at encoding is enough to prevent CP. Although previous findings that verbal interference in the ISI eliminated CP (Roberson & Davidoff, 2000) were interpreted as showing that verbal interference affected name retention, we suggest that it may, instead, have been due to interference with encoding. Verbal interference eliminates CP only if the interference type is *blocked* and the type of interference on each trial is known in advance. If interference type is randomized across trials, CP survives (Pilling et al., 2003). Knowing that verbal interference was to follow, the subjects may have decided there was no point in using a verbal strategy and did not name (encode) the target. With randomized interference, there is a two in three chance that naming the target will be followed by no interference or by visual interference, and they may have decided that, with these odds, it was worth naming the target.

Note that despite all of the above, it could still be the case that the advantage of cross-category over within-category discrimination is based on a direct matching of labels. If this were the case, the stimuli at the response stage in a delayed discrimination task must also be named correctly, and CP should be susceptible to Stroop interference at test presentation. We tested this idea in Experiments 2A and 2B. When Stroop interference was presented concurrently with test presentation, CP survived in all three conditions. This suggests that generation of the test stimuli label is not necessary for CP. The clear dissociation between Stroop interference at target presentation and Stroop interference at test presentation is inconsistent with a matching-of-labels account of CP.

The second question addressed here was whether the use and efficiency of name codes depend on aspects of the task design. There is evidence to suggest that the visual color code declines after approximately 1,500 msec (Pilling, 2001; see also Posner & Keele, 1967, on noncolor stimuli), and thus the reliance on verbal codes may be greater in tasks with long ISIs. This is consistent with some early work on Stroop effects. Flowers and colleagues (Flowers & Blair, 1976; Flowers & Dutch, 1976) have argued that tasks with low memory loads, which can be done on the basis of quick perceptual decisions, should not be susceptible to Stroop interference. Stimuli presentation times and the length of the ISI were reduced in Experiment 1B and subsequent experiments (relative to Experiment 1A), to test this. The pattern of results for discrimination accuracy found in Experiment 1B paralleled those in Experiment 1A, suggesting that verbal codes facilitated CP even when the memory load was substantially reduced. However, CP was not affected by Stroop interference in Experiments 2A and 2B (interference at response with a short ISI). This suggests that no access to the verbal codes of the test stimuli was required and is consistent with the perceptual or category code's surviving the short ISI and leading to CP.

However, it is possible that there was no effect of Stroop interference in Experiment 2A due to the relatively short (150-msec) presentation times. With two stimuli to process (the target and the foil), there may not have been enough time for the Stroop words to interfere with the naming of the colors. This is unlikely, since Stroop interference is found for durations as short as 30 msec (e.g., Shalev & Algom, 2000). However, to rule out this possibility, we increased the test stimuli presentation to 300 msec in Experiment 2B, thereby matching the Stroop exposure per stimulus to that in Experiments 1A and 1B. This did not significantly change the pattern of results; CP was still found across all the Stroop conditions. Since we found no evidence of any facilitation or interference, we suggest that the test stimuli in this task were not named and that color CP is not the result of direct matching of target and test labels.

If CP is not the result of a simple matching-of-labels process but target name generation is necessary, what leads to CP? We would argue that the function of target labeling is, in part, to activate a category code, which is then available to be used as part of the choice process (Bornstein & Korda, 1984; Pilling et al., 2003). This idea is consistent with evidence from the priming literature. Rosch (1975) and Neumann and D'Agostino (1981) have suggested that priming with a color word in color discrimination tasks leads to a mental representation that is visual in nature, but not as specific as the actual physical color code. Here, we argue that the category code could be activated either *top down* through labeling or *bottom up* by the color (the category exemplar). How the category code is activated is likely to depend on aspects of the task, such as the memory load. The Stroop word at target presentation (Experiments 1A and 1B) could have acted in a similar fashion as the word prime and, so, contributed to a top-down activation of the category code. In Experiments 2A and 2B, in which there were no Stroop words at target and little time to generate the verbal label (the target was presented only for 150 msec), the category code would have relied more on a bottom-up activation by the actual color. Our findings suggest that this category code was a reliable memory code when the memory demands were relatively low.

Other possible explanations for our findings have to be considered. First, it could be argued that the effect of incongruent Stroop interference in Experiments 1A and 1B was not a real Stroop effect. Adding an incongruent Stroop word to the color discrimination may simply have made the task more difficult, resulting in a reduction of CP due to generalized slowing of target processing. However, we think this is unlikely for two reasons. A generalized slowing account would not predict (and could not explain) the strong facilitation observed for cross-category comparisons in the congruent Stroop condition. Furthermore, if incongruent Stroop led to a general slowing of target processing, this should also have selectively affected within-category discrimination accuracy. This, however, was not the case.

Second, although we think that the incongruent color label in Experiments 1A and 1B reduced CP by preventing correct target labeling, it is also possible that CP was lost because the Stroop word was in direct conflict with the

target color. Since we used only blue and green stimuli and the corresponding color labels (BLUE, GREEN), the incongruent target label was always the correct foil label. If, for example, the target on an incongruent cross-category trial was blue with the word GREEN superimposed, the color label (GREEN) was correct for the foil presented alongside the target when a response had to be made. This may have supported a competing response—that is, to select the green (distractor) stimulus. Future studies could resolve this ambiguity by using other color labels (e.g., YELLOW, ORANGE) that do not match the actual colors of the target and foil stimuli. If our interpretation is correct, these kinds of color labels should lead to a reduction of CP in the same way as is shown here, since target name generation should be similarly affected. If, however, CP survived this manipulation, our model would have to be revised.

It is also important to acknowledge some potential problems that come with introducing Stroop interference to color discrimination tasks. Introducing Stroop interference at target presentation may have encouraged or even forced subjects to use a verbal strategy on the discrimination task. In other types of tasks and with other types of interference, it is possible that a more perceptual encoding strategy could be used, and that true CP (based on purely perceptual codes) could occur in such circumstances (see, e.g., Goldstone, Lippa, & Shiffrin, 2001; Harnad et al., 1991). What our data show is that even when verbal color labels are used at encoding, the advantage of cross-category over within-category discrimination is unlikely to be due to a direct comparison of verbal labels. Rather, we argue that the dual activation by target and target label produces a more durable or more reliable memory code and that it is this category code that facilitates comparisons with test colors and, so, contributes to the advantage of cross-category over within-category color discrimination.

#### AUTHOR NOTE

This research was supported by ESRC Studentship R42200034447 to A.J.W. The findings were first reported in A.J.W.'s thesis. We thank Amy Riddett, Ally Clifford, and Jessica Fielding for helping with the data collection, and we are grateful to Michael Pilling for his help in programming the experiments. We also thank Emre Özgen, Jules Davidoff, and one anonymous reviewer for their constructive comments. Correspondence concerning this article should be addressed to A. J. Wiggett, School of Psychology, University of Wales, Bangor, Gwynedd LL57 2AS, Wales (e-mail: a.wiggett@bangor.ac.uk).

#### REFERENCES

- BEALE, J. M., & KEIL, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217-239.
- BORNSTEIN, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287-300). Cambridge: Cambridge University Press.
- BORNSTEIN, M. H., & KORDA, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, *46*, 207-222.
- BOYNTON, R. M., FARGO, L., OLSON, C. X., & SMALLMAN, H. S. (1989). Category effects in color memory. *Color Research & Application*, *14*, 229-234.
- BRADLYN, A. S., & ROLLINS, H. A. (1980). Incidental memory for the color-word association in the Stroop color-word test. *Bulletin of the Psychonomic Society*, *16*, 269-272.

- CALDER, A. J., YOUNG, A. W., PERRETT, D. I., ETCOFF, N. L., & ROWLAND, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, **3**, 81-118.
- CARAMAZZA, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, **14**, 177-208.
- ETCOFF, N. L., & MAGEE, J. J. (1992). Categorical perception of facial expressions. *Cognition*, **44**, 227-240.
- FLETCHER, R. (1980). *The City University Color Vision Test* (2nd ed.). London: Keeler.
- FLOWERS, J. H. (1975). "Sensory" interference in a word-color matching task. *Perception & Psychophysics*, **18**, 37-43.
- FLOWERS, J. H., & BLAIR, B. (1976). Verbal interference with visual classification: Optimal processing and experimental design. *Bulletin of the Psychonomic Society*, **7**, 260-262.
- FLOWERS, J. H., & DUTCH, S. (1976). The use of visual and name codes in scanning and classifying colors. *Memory & Cognition*, **4**, 384-390.
- GILBERT, A. L., REGIER, T., KAY, P., & IVRY, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, **103**, 489-494.
- GOLDSTONE, R. L., LIPPA, Y., & SHIFFRIN, R. M. (2001). Altering object representations through category learning. *Cognition*, **78**, 27-43.
- HARNAD, S. (1987). Category induction and representation. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535-565). Cambridge: Cambridge University Press.
- HARNAD, S., HANSON, S. J., & LUBIN, J. (1991). Categorical perception and the evolution of supervised learning in neural nets. In D. W. Powers & L. Reeker (Eds.), *Working papers of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology* (Document D91-09, pp. 65-74). Kaiserslautern: Deutsches Forschungszentrum für künstliche Intelligenz.
- HUNT, R. W. G. (1987). *Measuring colour*. Chichester, U.K.: Wiley.
- KAY, P., & KEMPTON, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, **86**, 65-79.
- LAWS, G., DAVIES, I., & ANDREWS, C. (1995). Linguistic structure and non-linguistic cognition: English and Russian blues compared. *Language & Cognitive Processes*, **10**, 59-94.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 358-368.
- MACLEOD, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.
- NEUMANN, K. M., & D'AGOSTINO, P. R. (1981). Specificity of mental color codes. *American Journal of Psychology*, **94**, 451-459.
- NEWELL, F. N., & BÜLTHOFF, H. H. (2002). Categorical perception of familiar objects. *Cognition*, **85**, 113-143.
- ÖZGEN, E., & DAVIES, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, **131**, 477-493.
- PILLING, M. (2001). *On colour categorisation*. Unpublished doctoral thesis, University of Surrey.
- PILLING, M., WIGGETT, A., ÖZGEN, E., & DAVIES, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition*, **31**, 538-551.
- PISONI, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, **13**, 253-260.
- PISONI, D. B. (1991). Comment: Modes of processing speech and nonspeech signals. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman* (pp. 225-238). Hillsdale, NJ: Erlbaum.
- POSNER, M. I., & KEELE, S. W. (1967). Decay of visual information from a single letter. *Science*, **158**, 137-139.
- ROBERSON, D., & DAVIDOFF, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, **28**, 977-986.
- ROSCH, E. (1975). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 303-322.
- ROSCH, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- SCHYNS, P. G., & RODET, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 681-696.
- SHALEV, L., & ALGOM, D. (2000). Stroop and Garner effects in and out of Posner's beam: Reconciling two conceptions of selective attention. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 997-1017.
- SNOWDEN, C. T. (1987). A naturalistic view of categorical perception. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 332-354). Cambridge: Cambridge University Press.
- STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.
- THOMPSON, E. (1995). *Colour vision: A study in cognitive science and the philosophy of perception*. London: Routledge.
- UCHIKAWA, K., & SHINODA, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research & Application*, **21**, 430-439.
- WINAWER, J., WITTHOFT, N., FRANK, M. C., WU, L., WADE, A. R., & BORODITSKY, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, **104**, 7780-7785.
- YOUNG, A. W., ROWLAND, D., CALDER, A. J., ETCOFF, N. L., SETH, A., & PERRETT, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, **63**, 271-313.

# **APPENDIX** **ANCOVA Results: Accuracy With Median Reaction Times** **for All Experiments**

Experiment 1A	Category	$F(1,26) = 9.32, MS_e = 74.55, p < .01$
	Interference	$F(2,53) = 7.68, MS_e = 47.36, p < .001$
	Category $\times$ interference	$F(2,53) = 14.09, MS_e = 55.11, p < .001$
Experiment 1B	Category	$F(1,19) = 25.09, MS_e = 54.57, p < .001$
	Interference	$F(2,39) = 1.99, MS_e = 42.86, p > .05$
	Category $\times$ interference	$F(2,39) = 11.96, MS_e = 62.93, p < .001$
Experiment 2A	Category	$F(1,26) = 20.31, MS_e = 160.50, p < .001$
	Interference	$F(2,53) = 0.06, MS_e = 51.36, p > .05$
	Category $\times$ interference	$F(2,53) = 0.06, MS_e = 58.81, p > .05$
Experiment 2B	Category	$F(1,22) = 26.74, MS_e = 74.93, p < .001$
	Interference	$F(2,45) = 8.84, MS_e = 46.48, p < .001$
	Category $\times$ interference	$F(2,45) = 1.65, MS_e = 47.60, p > .05$