Is color “categorical perception” really perceptual?

MICHAEL PILLING, ALISON WIGGETT, EMRE ÖZGEN, and IAN R. L. DAVIES
University of Surrey, Guildford, England

Roberson and Davidoff (2000) found that color *categorical perception* (CP; better cross-category than within-category discrimination) was eliminated by verbal, but not by visual, interference presented during the interstimulus interval (ISI) of a discrimination task. On the basis of this finding, Roberson and Davidoff concluded that CP was mediated by verbal labels, and not by perceptual mechanisms, as is generally assumed. Experiment 1 replicated their results. However, it was found that if the interference type was uncertain on each trial (Experiment 2), CP then survived verbal interference. Moreover, it was found that the target color name could be retained across the ISI even with verbal interference (Experiment 3). We therefore conclude that color CP may indeed involve verbal labeling but that verbal interference does not necessarily prevent it.

Categorical perception (CP) is prevalent throughout perception. Its signature is better discrimination across category boundaries than for equivalently separated stimuli within the same category (Harnad, 1987). Despite its name, it is unclear whether CP is truly a perceptual phenomenon. Here, we investigate Roberson and Davidoff's (2000) claim that CP arises from the retention of verbal labels in working memory, rather than from perceptual processes. We also consider the suggestion that CP is due to the *shift-toward-prototype* effect in memory: Recall of category exemplars is biased toward the prototype (Huttenlocher, Hedges, & Vevea, 2000). We will continue to use the term CP for the moment, following conventional usage, despite questions over its appropriateness, but this implies no theoretical position.

The first demonstrations of CP were in speech perception. Given a continuum of sounds varying from one phoneme to another in equal steps (e.g., /ba/ to /da/), sounds straddling a phonemic category boundary are more discriminable than sounds within the category (e.g., Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1991). Early theorists stressed the uniqueness of the

speech domain in CP, claiming that speech perception was affected by speech production—the so-called motor theory of CP (Liberman et al., 1957).

Subsequent research showed that CP is a general characteristic of perception, including vision. For instance, CP has been found in the perception of line length (Tajfel & Wilkes, 1963), facial expressions (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Young et al., 1997), facial actions (Campbell, Woll, Benson, & Wallace, 1999), faces of different species (Campbell, Pascalis, Coleman, Wallace, & Benson, 1997), and familiar objects (Newell & Bühlhoff, 2002). Categorical effects also occur in color perception across a range of tasks, including recognition memory (Uchikawa & Shinoda, 1996), same–different judgments (Bornstein & Korda, 1984; Boynton, Fargo, Olson, & Smallman, 1989), and similarity judgments (Laws, Davies, & Andrews, 1995). For instance, Bornstein and Korda (1984) examined discrimination for blue and green colors spaced equally in Munsell hue steps (2.5 steps). Color pairs were presented in a delayed *same–different* discrimination task with a 250-msec interstimulus interval (ISI). Discrimination was better for colors straddling the blue–green category boundary than for colors within the same category.

Apart from the motor theory, three main classes of theory have emerged from the literature: naturalistic theories, perceptual change theories, and labeling theories. Naturalistic theories (e.g., Bornstein, 1987; Snowden, 1987) regard CP as 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. These theories therefore assume that CP is both inborn and perceptually mediated. Naturalistic theories are supported by CP in young human infants and in animals. In color perception, infants perceive colors categorically before they

This research was supported in part by ESRC Studentships R00429824622 and R42200034447 to M.P. and A.W., and in part by ESRC Project Grant R000238310 to I.R.L.D., Davidoff, and Corbett. Experiments 1A and 1B were first reported in M.P.'s thesis, and Experiments 2A and 2B will form part of A.W.'s thesis. They have also been briefly reported at the conference of the Cognitive Section of the British Psychological Society (2001). We are very grateful to Amy Riddett and Ally Clifford for collecting the naming data and to John Groeger for discussions of working memory and his suggestions of appropriate interference tasks. We also thank the three reviewers for their constructive comments on an earlier version. We particularly thank Dale Klopfer for suggesting the shift-toward-prototype account as an alternative explanation of the categorical perception effect. Correspondence concerning this article should be sent to I. R. L. Davies, Department of Psychology, University of Surrey, Guildford GU2 7XH, England (e-mail: i.davies@surrey.ac.uk).

learn names for them (Bornstein, Kessen, & Weiskopf, 1976; cf. Davies & Franklin, 2002; Franklin & Davies, 2002; Gerhardstein, Renner, & Rovee-Collier, 1999). Color CP has also been found in several nonhuman species, including pigeons and macaque monkeys (Bornstein, 1974; Sandell, Gross, & Bornstein, 1983; Walsh, Kulikowski, Butler, & Carden, 1989). However, naturalistic theories do not give a complete account of CP, particularly of its variability and malleability. For instance, Zatorre and Halpern (1979) found that CP of musical intervals was greater in trained musicians than in non-musicians. In color perception, CP differs for speakers of different languages (e.g., Kay & Kempton, 1984). Furthermore, CP can emerge from high-level contextual cues (Goldstone, 1995) or through the learning of new categories (Goldstone, 1994; Livingston, Andrews, & Harnad, 1998; Özgen & Davies, 2002).

Perceptual change theories (e.g., Harnad, Hanson, & Lubin, 1991; for the altered object description theory, see Goldstone, Lippa, & Shiffrin, 2001) propose that learning to categorize a stimulus dimension, either through massed practice or through learning to distinguish stimuli by name, results in representational change of the stimulus dimension. This perceptual change could be in the sensitivity around boundaries (e.g., Harnad et al., 1991) or in the creation of new perceptual structures (e.g., Schyns & Rodet, 1997). Such theories are similar to naturalistic theories in assuming that CP effects are perceptually mediated but differ in considering that the effects are acquired rather than inherent. Evidence of the malleability of low-level visual processes (perceptual learning) supports this position (see Goldstone, 1998, for a review). For instance, massed discrimination practice improves perceptual performance, but such improvements often show stimulus and retinotopic specificity, suggesting an involvement of early visual processing (Ahissar & Hochstein, 1997; Fahle, 1994; Fahle, Edelman, & Poggio, 1995). However, although category-training experiments have documented acquisition of CP effects in the short term (e.g., Goldstone, 1994), no evidence of a change in baseline sensitivities or of any long-term change has been reported.

Labeling, or direct language, theories (e.g., Fujisaka & Kawashima, 1971; Kay & Kempton, 1984; Roberson & Davidoff, 2000; Rosen & Howell, 1987) propose that discrimination decisions are based, to some degree, on comparisons of verbal labels of the stimuli. Stimuli that straddle category boundaries have different labels, whereas stimuli from the same category have the same label. Labeling thus helps cross-category judgments, but not within-category judgments. To illustrate, consider a same-different task comparing within and cross-category discrimination that uses four possible stimuli from two adjacent categories (blue and green): blue1 blue2 | green2 green1. These are ordered from blue1 to green2, so that blue1 is more similar to blue2 than it is to green2, and so on. If judgments are based on naming alone and naming is completely reliable, cross-category pairs (blue2–

green2) would be named differently and should be judged as different with complete accuracy. Within-category pairs (blue1–blue2 and green1–green2) would always be given the same name and should be judged as the same, erroneously, on all trials. Same trials should be error free. This would give better cross-category than within-category discrimination, characteristic of CP. More realistically, naming reliability is likely to be less than perfect, particularly for stimuli close to category boundaries (blue2 and green2). Nevertheless, cross-category judgments should still be more accurate than within-category judgments, under any realistic assumption about naming reliability.

More realistically still, judgments are likely to be based on comparison of labels and comparison of perceptual representations. Within-category judgments may be based primarily on comparison of perceptual codes, whereas cross-category judgments may be based on comparisons of perceptual codes and verbal labels. We will later use a simple direct-naming model to derive upper and lower bounds for judgments based on naming alone. If observed accuracy is within these bounds, it would be evidence consistent with a pure labeling account. If observed accuracy exceeds these predictions, this would be evidence for some involvement of perceptual representations in decisions.

There are some difficulties for straightforward labeling accounts of CP. Identification boundaries do not always align precisely with peaks in discrimination (Calder et al., 1996). Differences in naming do not always lead to differences in CP (Malt, Sloman, Gennari, Shi, & Wang, 1999). Furthermore, Goldstone et al. (2001) found evidence against labeling accounts of CP acquired through categorization training. Perceived similarity between two faces in the same trained category and a third neutral stimulus became more similar after category training. Goldstone et al. argued that labeling cannot account for this change, because the neutral stimulus was never seen during training and, therefore, could not have been labeled.

In contrast, Roberson and Davidoff (2000) found compelling evidence in favor of a direct language account of CP. They investigated whether color CP was perceptually mediated or verbally mediated. They used a delayed two-alternative forced-choice (2AFC) procedure with a 5-sec ISI between presentations of the target color and the two test colors. Stimuli were either blue or green, and test pairs either crossed the blue–green boundary (cross-category pairs: blue2–green2) or were from the same category (within-category pairs: blue1–blue2 or green1–green2). Perceptual distances within test pairs were the same in both cases (in Munsell metric; see Hunt, 1987; Munsell, 1923). One color was presented as the target (e.g., blue2), and the test pair consisted of the target and either a same-category distractor (blue1) or a different-category distractor (green2). The task was to decide, as indicated by a keypress, which color in the test pair was physically identical to the target.

During the ISI, one of three kinds of interference occurred: no interference, visual interference, or verbal in-

terference. In the no-interference condition, there was a blank screen throughout the ISI. In the visual interference condition, there was a stationary curved line on the screen for the duration of the ISI, and it had to be tracked visually. In the verbal interference condition, a list of words was presented on the screen, and these had to be read aloud for the duration of the ISI. Note that performance on the interference tasks was not recorded.

In the no-interference condition, a classic CP effect was found: Discrimination was more accurate for cross-category pairs than for within-category pairs. Visual interference reduced accuracy for both pair types, but the CP effect persisted. Verbal interference also reduced accuracy, but this effect was greater for cross-category pairs than for within-category pairs, eliminating the CP effect. It made no difference whether the interfering words were nonbasic color words (Experiment 1: e.g., navy, khaki, teal) or noncolor words (Experiment 3: e.g., foil, lash, tact). Roberson and Davidoff (2000) argued that verbal interference selectively interfered with retention of the categorical representation of the target color. Since visual interference did not affect CP, they concluded that the category code must be verbal and that CP was not a perceptual phenomenon.

Huttenlocher et al. (2000) also argued that CP is, in essence, a phenomenon of memory, rather than of perception. However, rather than arising from verbal labeling, they argued that CP was due to bias toward the category prototype during recall (see also Bartlett, 1932). Consider the four stimuli used in previous examples plus their category prototypes (P): P_{blue} blue1 blue2 green2 green1 P_{green} . The category prototypes are further away from the boundary than any of the experimental stimuli (see, e.g., Rosch, 1972). Following presentation of a target, its memory representation should shift toward the prototype during the retention interval (ISI). On a cross-category trial (blue2↔green2; the double arrow indicates either order), the memory representation of the target should shift away from the category boundary, thus increasing its separation from the representation of the test stimulus. On the other hand, for the same target on a within-category trial, the same shift toward the prototype will reduce its separation from the representation of the test stimulus (blue1 or green1). These differential effects—compression or expansion in representation space—would produce more accurate cross-category judgments than within-category judgments, or apparent CP. However, there is a further prediction for within-category trials: The effect of the shift toward the prototype should depend on stimulus order. For example, if blue1 is the target, its shift toward the prototype *increases* its separation from the test stimulus (blue2), whereas if blue2 is the target, its shift toward the prototype *decreases* its separation from the test stimulus (blue1). Hence, if CP is based on a shift toward a prototype, accuracy for within-category trials should depend on the stimulus sequence. We tested this prediction in our first two sets of experiments.

To provide a comparison with the prototype bias account, we also derived quantitative predictions for cross-category and within-category judgments from a simple labeling model. We assumed that judgments are based on naming alone and that each stimulus is named either *blue* or *green*. Furthermore, it was assumed that for the 2AFC task, if the target and one of the test stimuli are given the same name and this is different from the name given to the other test stimulus, the test stimulus with the same name will be chosen. If both test stimuli are given the same name, one will be chosen randomly. This model will be developed further below, together with its equivalent for same–different judgments. Predictions from the model provided an upper bound for accuracy based on naming alone and was used to probe the adequacy of the direct-naming account.

Finally, note that because sequential stimulus presentation is commonly used (e.g., Campbell et al., 1997; Livingston et al., 1998), this seems to implicate the involvement of memory processes in performance. However, simply because there is a memory component to many tests of CP does not necessarily imply that CP arises from memory processes. The origin of CP could be in the perceptual representations of the stimuli that are read into memory, in the way memory deals with these, or both. We used sequential presentation in our experiments to replicate Roberson and Davidoff's (2000) methods.

Rationale and Outline of the Experiments

The experiments reported here were designed primarily to test Roberson and Davidoff's (2000) direct language account of CP. In addition, they were also used to test the prediction from Huttenlocher et al.'s (2000) shift-toward-prototype account of CP and to test the predictions from the simple direct-naming model. The logic of the experiments required stimuli that were named predominantly either *blue* or *green* and whose separations were equated in a perceptually uniform metric. The latter was achieved using the CIE (1976) Lu^*v^* color space (see Hunt, 1987; Wyszecki & Stiles, 1982), and the categorical status of the stimuli was checked in a preliminary naming experiment. The data from the latter were also needed for the direct-naming model.

The first two experiments were run as 2AFC (Experiments 1A and 2A) and as *same–different* (Experiments 1B and 2B) discrimination tasks. These methods are often regarded as equivalent. However, because the 2AFC method uses two test stimuli, whereas the *same–different* method uses one test stimulus, the former may encourage a discriminative response, whereas the latter may invoke a categorical response (see Macmillan, 1987, pp. 57–58). If this is the case, CP may survive verbal interference with the *same–different* method, whereas it may be eliminated with the 2AFC method. These variations also served to test the generality and robustness of the findings.

Experiment 1 was a partial replication of Roberson and Davidoff's (2000) first two experiments but used

more carefully controlled interference tasks. Their visual interference task required no overt response, whereas their verbal interference task did. In addition, performance was not measured on either task. Thus, either because of implicit differences in the demand characteristics of the two tasks or because of inherent differences in complexity, the different effects of the tasks may have been due to differences in effective cognitive load, rather than to modality difference per se. Experiment 2 tested the possibility that Roberson and Davidoff's key result—verbal interference eliminated CP—arose from an aspect of their procedure. Their interference conditions were blocked (as were ours in Experiment 1) and, therefore, were predictable. This may have led to the use of different strategies in different interference conditions to try to avoid the effects of interference. In Experiment 2, therefore, the interference trials were randomly interleaved, rather than blocked. Experiment 3 tested an implicit assumption of the direct-naming account, that verbal interference hinders the retention of a color name over the ISI, whereas visual interference does not. A target color was presented, followed by visual, verbal, or no interference, followed by a test for the name of the target. If verbal interference blocked name retention, performance in this condition should be particularly affected. We also use the results of the first two experiments to test the shift-toward-prototype prediction for the within-category order effect described earlier and to compare within- and cross-category performance with predictions from a simple direct-naming model.

EXPERIMENT 1

Replication (1A, 2AFC; 1B, *Same-Different*)

Although Roberson and Davidoff (2000) found a reliable effect of verbal interference on CP, there was a possible artifact in their method. Whereas the verbal interference task required an overt response, the visual interference task required only scanning eye movements that were not monitored. This may have resulted in the visual task's being attended to less than the verbal task, resulting in the differential effect of verbal interference on CP. Consistent with this possibility, visual interference affected overall accuracy less than did verbal interference (see particularly their Experiment 2).

Experiment 1A used the same general method as the first two experiments in Roberson and Davidoff (2000) but used modified interference tasks. The method was a 2AFC delayed discrimination task with no interference, visual interference, or verbal interference during the ISI. The interference tasks were intended to be treated as equally important, and both required the same overt response. Responses on these interference tasks were recorded, as well as responses on the primary task, so that both could be used in the analysis. The choice of interference tasks was motivated by the theory of working memory (Baddeley & Hitch, 1974). The visual interference task required judging whether the first of two

shapes presented sequentially could fit inside the second shape and required the use of the visual subsystem of working memory. The verbal interference task required a rhyme judgment to be made. The rhyming words were orthographically dissimilar (e.g., *yacht* and *plot*) and, therefore, required phonological comparison to determine whether they rhymed (Brown, 1987). As was suggested earlier, the 2AFC method is perhaps more likely to invite a judgment of relative identity, whereas a *same-different* task may be more likely to encourage a judgment of absolute identity. Thus, CP effects in *same-different* judgments may be stronger and less vulnerable to verbal interference during the ISI than those in a 2AFC task. Therefore, to test the generality of Roberson and Davidoff's results, Experiment 1A was repeated as a delayed *same-different* judgment task (Experiment 1B).

Method

Participants

Twenty people (10 men and 10 women; age range, 18–22 years) took part in the preliminary naming study, and 28 people (3 men and 25 women; age range, 18–26 years) took part in each main experiment. All were students at the University of Surrey; some received course credit for participation.

Stimuli and Apparatus

All the stimuli were displayed on a 15-in. Sony Trinitron SVGA monitor controlled by a 100-MHz 80486 IBM PC compatible computer fitted with a Diamond Stealth graphics card. Control of the experiments and randomization of conditions were done through purpose-made programs written in Microsoft Visual Basic (v. 5.0). Special software functions were used to provide accurate event timing. All the experiments took place in a darkened room. Emulation of Munsell colors was supported by software functions (Color Science Library v. 2.0) obtained from the Computer Graphics Systems Development Corporation (Mountain View, CA). Calibration of the monitor was performed according to the procedure in Travis (1991). The participants were tested for color vision problems, using the City University color vision test (Fletcher, 1980).

According to Bornstein and Monroe (1980), the *blue-green* category boundary for English speakers is at Munsell hue 7.5 BG. Stimuli were selected so that half fell to one side of the boundary and half fell on the other side. Naming was also checked in a preliminary study to confirm their suitability (see the Results section). There were 16 color stimuli of constant Munsell chroma (saturation) 7.4, but varying in hue and value (lightness). Four hues (1.44 B, 8.76 BG, 6.06 BG, and 4.04 BG) were combined with four values (5.40, 5.80, 6.20, and 6.60) to give a matrix of 16 colors. The perceptual distance between adjacent hues was 9 units in CIE (1976) Lu^*v^* space, and the perceptual distance between adjacent values was 4 units. The chromaticity of all the color stimuli was checked using a Minolta CS-100 Colorimeter. Each color stimulus was 55 mm², subtending 6.3° of visual arc at an approximate viewing distance of 500 mm.

For Experiment 1B, the same stimuli as those used in Experiment 1A were used as different pairs. However, in accord with the *same-different* method, identical pairs were presented, as well as different pairs. Each of the 16 stimuli was used as an identical pair.

For the visual interference task, 72 shape pairs were used. All the shapes were two-dimensional outline drawings of closed abstract shapes in black lines on a neutral gray background. The shapes in each pair were always physically different from each other. For half of the shape pairs, the first shape could fit in the second shape; for the other half, it could not. For the verbal interference task, 72 word

pairs were used. Some of these word pairs were taken from Rack (1985) and Brown (1987). For half of the pairs, the two words rhymed, and for half they did not. The word pairs were displayed on the computer monitor in Arial font on a neutral background.

Procedure

Naming. There were two tasks: forced choice (blue or green) and free responses. The order was counterbalanced across participants. Each stimulus was presented individually until a response was made. In the forced-choice version, this was indicated by a key-press; in the free-response method, responses were spoken aloud and were recorded by the experimenter.

Experiment 1A: 2AFC. All adjacent hue pairs in the 4×4 matrix were used, giving eight within-category pairs and four cross-category pairs. Stimulus pairs were chosen at random for each trial, but all the pairs were shown with equal frequency in all the conditions. Each stimulus was shown as a target and as a distractor with equal frequency. On each trial, the target was shown for 1,000 msec, followed by a 5,000-msec filled or unfilled ISI, and then two test stimuli (target plus distractor) were shown for 1,000 msec or until a response was made. The position of the target (left or right) was randomized, with the constraint that it appeared in each position on half of the trials. The intertrial interval was 600 msec. Instructions emphasized accuracy for the primary task and both speed and accuracy for the secondary tasks. There were six blocks of 36 trials, two for each type of interference, with a short break between blocks. Interference type was randomized across blocks. The participants were informed of the type of interference before each block of trials.

In the no-interference condition, there was a neutral gray screen for the duration of the ISI. In the visual interference condition, one of the 72 shape pairs was randomly selected and presented during the ISI. Each shape pair was shown only once. The first shape was shown 300 msec after the offset of the target and was presented for 1,500 msec. This was followed by a blank 1,000-msec interval; then the second shape was shown for 1,500 msec. The participants responded by pressing one of the arrow cursor keys on the keyboard (left arrow, the shape fits; right arrow, the shape does not fit). In the verbal interference condition, 1 of the 72 word pairs was randomly selected and presented. Each word pair was shown only once. The sequence of presentation and durations of the stimuli were the same as those for the shape pairs, as were the response keys.

For both interference conditions, responses were required before the offset of the second stimulus (word or shape), to avoid overlap between responses to the interference tasks and presentation of the test colors. Late responses were recorded as misses. Response times for the verbal and visual interference tasks were measured from onset of the second stimulus to the response.

For each condition, there was a block of 10 practice trials before the experimental trials. Color pairs were randomly selected for these trials, but all the word and shape pairs used in the practice trials were not used in the experimental trials. Auditory feedback was given on the practice trials to indicate errors on the 2AFC task or the interference task. No feedback was given in the experimental trials.

Experiment 1B: Same-Different. The procedure was the same as that for Experiment 1A, except in those ways required to implement the *same-different* method. A single target was presented, followed by a single probe, and the required judgment was whether they were physically identical (same) or different. Decisions were indicated by a mouse buttonpress: left, same; right, different. As with the 2AFC task, the participants were instructed to make speeded responses only on the interference tasks, and not on the *same-different* task. Each of the 16 identical pairs was presented with equal frequency in each condition. To have the same number of trials and maintain the same ratio of cross-category and within-category trials as in Experiment 1A and to have an equal number of

Table 1
Naming Frequencies (%) for Forced Choice (Blue or Green) and Free-Response Conditions

Condition	Blue		Green	
	Boundary	Nonboundary	Boundary	Nonboundary
Forced choice	85.60	96.87	81.88	98.75
Free response	66.57	92.50	76.25	95.94

trials of each of the 16 physically identical pairs, a ratio of 2:1 *same* and *different* trials was used in each condition.

Results

Naming

Naming frequencies across respondents for the four hues averaged across lightness are given in Table 1. All the stimuli with hue 1.44 B and 8.76 BG had the dominant name *blue*, and all the stimuli with hue 6.06 BG and 3.46 BG had the dominant name *green*. Respondents were less consistent in naming the boundary stimuli than the stimuli nearer the category centers. Naming reliability was also slightly lower in the free-response condition. Even so, the lowest naming frequency was 66.57%, for the boundary blue.

Experiment 1A: 2AFC

Mean accuracy was 75.40% and 84.23% for the visual and the verbal interference tasks, respectively. Median correct response times on the secondary tasks were calculated for each participant. Mean response times across participants were 920 msec for visual interference and 890 msec for verbal interference.

Mean accuracy on the primary task in each condition for cross-category and within-category color pairs is shown in Figure 1, panel A. The results were analyzed with a two-way analysis of variance (ANOVA) with pair type (cross or within) and interference type (none, visual, or verbal) as factors. The ANOVA was followed by post hoc protected *t* tests. Both main effects and the interaction were significant. Accuracy was higher for cross-category pairs than for within-category pairs [$F(1,27) = 6.40$, $MS_e = 92.08$, $p < .05$]. The significant interference effect [$F(2,54) = 17.16$, $MS_e = 60.29$, $p < .0001$] reflected lower mean accuracy for both interference conditions, as compared with a baseline, for both within-category and cross-category conditions [minimum difference, $t(25) = 2.32$, $p < .05$]. The significant interaction [$F(2,54) = 3.47$, $MS_e = 53.55$, $p < .05$] occurred because there was an advantage for cross-category pairs over within-category pairs in the no-interference condition [$t(26) = 3.16$, $p < .01$] and in the visual interference condition [$t(26) = 2.82$, $p < .01$], but not in the verbal interference condition [$t(26) = 0.23$, $p > .05$]. For the cross-category condition, mean verbal interference was less than mean visual interference, whereas the reverse was the case for within-category stimuli, although neither difference reached significance: [$t(25) = 1.98$, $p =$

Figure 1. Panels A and B show accuracy in Experiments 1A and 1B for cross- and within-category pairs. Panel A shows Experiment 1A (two-alternative forced choice). Panel B shows Experiment 1B (*same-different*). Panels C and D show accuracy in Experiments 1A and 1B for within-category pairs when the boundary stimulus was the target and when it was the test. Panel C shows Experiment 1A. Panel D shows Experiment 1B. Error bars show ± 1 standard error in all panels.

.059, and $t(25) = 1.67$, $p > .05$, for cross-category and within-category, respectively].

Experiment 1B: Same-Different Judgments

Individual accuracy scores on the primary task were calculated using the A' statistic (see, e.g., Swets, 1996). A' combines hits and false positives and provides a measure of discrimination that is independent of response bias. It varies between .5 (chance) and 1 (perfect discrimination). Mean scores are shown in Figure 1, panel B.

The same analysis as that in Experiment 1A was performed for Experiment 1B. Both main effects were significant. Accuracy was significantly better for cross-category pairs than for within-category pairs [$F(1,27) = 6.54$, $MS_e = 0.098$, $p < .05$], there was a significant effect of interference [$F(2,54) = 3.90$, $MS_e = 0.012$, $p < .05$], and the pair type $\times$ interference interaction was also significant [$F(2,54) = 3.20$, $MS_e = 0.044$, $p < .05$]. Post hoc testing revealed the same significant effects as those in Experiment 1.

The selective effect of verbal interference on CP shown in Experiments 1A and 1B could be due to different levels of attention or effort given to the two interference tasks. This possibility was tested in two ways. First, the

results of both experiments were reanalyzed using two-way analyses of covariance (pair type $\times$ interference type), with median reaction times on the interference task as a covariate. Second, the two-way ANOVAs were repeated using only trials on which a correct response on the secondary task had been given. Neither reanalysis altered the conclusions.

Shift Toward Prototype

Figure 1, panel C shows the mean accuracy scores on the 2AFC task for the within-category judgments, broken down by whether the boundary stimulus was the target (presented first) or the nontarget (presented second) of the two test stimuli. It can be seen that there was clearly no effect of order for the no-interference condition or for the verbal interference condition. There was a small but nonsignificant difference in the predicted direction for the visual interference condition: Accuracy was lower when the boundary stimulus was the target than when the boundary stimulus was the test.

Figure 1, panel D shows the equivalent data for the *same-different* task, but accuracy is indicated by A' . In this case, there was an order effect. Accuracy was greater for within-category trials when the boundary stimulus

occurred first (e.g., blue2→blue1) than when the order was reversed [e.g., blue1→blue2; $F(1,27) = 84.41$, $MS_e = 0.009$, $p < .0001$]. There was an effect of interference condition [$F(2,54) = 6.22$, $MS_e = 0.058$], but no interaction [$F(2,54) = 0.35$, $MS_e = 0.009$, $p > .05$].

Tests of Direct Naming

Here, we used the naming data from Table 1 as the best estimate of naming accuracy. The direct-naming model predicts accuracy, assuming that judgments are based on comparisons of color names alone and that every stimulus is named either *blue* or *green*. The rationale is slightly different for the two tasks, since the 2AFC task requires comparison of the target name with the names of both test stimuli, whereas the *same-different* task requires only a single comparison. For the 2AFC task, the model assumes that the target name is remembered across the ISI and is then compared with the names of the test stimuli generated when they are presented. If naming comparisons yield an unequivocal choice (target name is the same as one test stimulus name and different from the other), that choice will be made. If the target name matches both test stimulus names (an equivocal trial), a random choice will be made. The probability of a correct choice is then the sum of the probability of a correct choice on unequivocal trials plus half the probability of an equivocal trial.

The probability of a correct choice on an unequivocal trial is given by Formula 1:

$$[p(T-N1) \times p(T-N1) \times p(D-N2)] + [p(T-N2) \times p(T-N2) \times p(D-N1)], \quad (1)$$

where N1 and N2 are the two possible names, the target is T, the distractor (incorrect test stimulus) is D, and $p(T-N1)$ stands for the probability that the target is named N1, and so on. Note that a correct choice can be made either by naming all of the stimuli correctly (first half of the formula or by naming all of the stimuli incorrectly (second half of the formula). The probability of a correct choice on an equivocal trial (all named the same) is given by Formula 2:

$$\{[p(T-N1) \times p(T-N1) \times p(D-N1)] + [p(T-N2) \times p(T-N2) \times p(D-N2)]\}/2. \quad (2)$$

The probability of a correct choice is then the sum of the results of the two formulae.

The process is simpler for the *same-different* task. If the target and the test stimuli names are the same, a *same* decision will be made; otherwise, a *different* decision will be made. The probability of a correct decision for same trials is given by

$$p[(T-N1) \times p(P-N1)] + [p(T-N2) \times p(P-N2)], \quad (3)$$

where P is the second, or probe, stimulus, and the probability of a correct choice on a different trial is given by

$$[p(T-N1) \times p(P-N2)] + [p(T-N2) \times p(P-N1)]. \quad (4)$$

Table 2
Predicted and Observed Accuracy Given as Percentage Correct (Experiment 1A, Two-Alternative Forced Choice) and A' (Experiment 1B, *Same-Different*) for the No-Interference Condition

Condition	Accuracy	
	Predicted	Observed
Experiment 1A		
Within-category, blue	57.00	71.27
Within-category, green	58.00	66.96
Cross-category	78.00	76.04
Experiment 1B		
Within-category, blue	0.52	0.69
Within-category, green	0.52	0.71
Cross-category	0.79	0.76

The predicted scores for every possible stimulus combination were calculated for both tasks by substituting the naming probabilities taken from the naming task. For the *same-different* task, the predicted probabilities were used to calculate A' . The predictions are shown in Table 2, together with the observed data for the no-interference condition. It can be seen that the cross-category prediction is very close to the observed data but that the within-category prediction is an underestimate. The results for the *same-different* task are shown in the bottom half of Table 2, and again, the prediction is close for the cross-category condition but is too low for the within-category condition.

To summarize the main results, there was a significant advantage in discriminating cross-category pairs over within-category pairs in the no-interference condition. This advantage for cross-category pairs persisted in the visual interference condition, although overall accuracy was lower. However, this advantage was lost in the verbal interference condition. There was no order effect for the within-category trials for the 2AFC task, but there was an effect for the *same-different* task. However, this effect was in a direction opposite to that predicted by the shift-toward-prototype account. Finally, a simple direct-naming model predicted the cross-category scores well for both the 2AFC and the *same-different* versions of the task.

Discussion

For both experiments, mean accuracy rates indicated above-chance discrimination of color pairs in all the conditions. With no interference, a CP effect was found. That is, despite the fact that the perceptual differences between all the color pairs were equal (in CIE units), discrimination of cross-category pairs was more accurate than that of within-category pairs. Both interference conditions resulted in a reduction in overall accuracy; however, the CP effect persisted with visual interference but was lost with verbal interference. Further analysis suggested that the effect of verbal interference was not related to performance on the interference tasks. The same effect of verbal interference was found, relative to

the visual interference condition, even when performance on the interference tasks was taken into account.

The size of the accuracy scores on cross-category trials was well predicted by a simple direct-naming model. However, the same model underestimated within-category performance. For the 2AFC task, this may be because, if both test stimuli are given the same name, the decision is based on comparison of visual representations of the test stimuli with the visual memory of the target. However, in the *same-different* task, if the target and the test are given the same name, there is nothing in the name information alone to distinguish within-category *different* trials from within-category *same* trials. If visual information influences the decision, as seems probable, this implies that the visual information is being processed anyway, rather than being turned to, when name information is insufficient to make the choice. This suggests that even in the 2AFC task, visual information may also be processed automatically.

The tests of the shift-toward-prototype account of CP did not support it. Accuracy on within-category trials was independent of the order in which the boundary and the nonboundary stimuli occurred for the 2AFC task and in the opposite direction for the *same-different* task. Neither direct-naming nor perceptual-warping accounts of CP predict within-category order effects. Moreover, although we speculated that 2AFC and *same-different* tasks may not be wholly equivalent, the possibility of differential order effects was not considered. In view of the unexpectedness of this result, we will defer further discussion until we see whether it was replicated in the next experiment.

EXPERIMENT 2

Unpredictable Interference (2A, 2AFC; 2B, *Same-Different*)

Experiments 1A and 1B replicated the effects reported in Roberson and Davidoff (2000). However, as in Roberson and Davidoff, interference conditions were ordered in blocks for both experiments. The participants were also forewarned of the interference type before each block and, therefore, were probably aware of the type of interference that they would encounter on each trial. This may have encouraged the use of different coding strategies in the various interference conditions. Verbal interference may have affected CP because it interfered with the retention of verbal codes or because it discouraged a verbal strategy.

In Experiments 2A and 2B, the order of interference conditions was randomized, and the participants were not forewarned of the type of interference that would occur on each trial. This could have two effects. The naming strategy could be abandoned altogether, which should eliminate CP in all conditions if the direct language account is true. Alternatively, the participants could try to use the name on all trials, and if the name could be retained, CP should be found in all conditions. If verbal interference blocked name retention, CP should be found only in the baseline and visual interference conditions.

Method

Twenty-eight participants took part in Experiment 2A (22 women and 6 men; age range, 18–42 years) and 26 participants in Experiment 2B (21 women and 5 men; age range, 18–39 years). All were students at the University of Surrey; some received course credit for their participation. None had taken part in the previous experiments. The stimuli, apparatus, and procedure were the same as those used in Experiment 1, except that the order of interference conditions was randomized within a block, with the constraint that they occurred with equal frequency.

Results

Experiment 2A: 2AFC

Data analysis followed the same scheme as that in Experiments 1A and 1B. Mean accuracy and response times on the interference tasks were 72.52% and 992 msec and 78.24% and 957 msec for visual and verbal interference, respectively. Mean accuracy scores on the primary task for cross-category and within-category pairs in Experiment 2A are shown in Figure 2, panel A. A significant main effect was found for pair type [$F(1,27) = 16.15$, $MS_e = 106.49$, $p < .01$] and interference condition [$F(2,54) = 26.85$, $MS_e = 56.70$, $p < .01$]. The pair type $\times$ interference interaction, however, was clearly not significant [$F(2,54) = 0.13$, $MS_e = 56.02$, $p > .05$].

Post hoc testing (protected t tests) showed that there was a statistically significant advantage for cross-category pairs over within-category pairs in the no-interference condition [$t(26) = 2.12$, $p < .05$]. This difference was also significant in both the visual interference condition [$t(26) = 2.61$, $p < .05$] and the verbal interference condition [$t(26) = 2.13$, $p < .05$]. For cross-category pairs alone, accuracy was significantly better in the no-interference condition than in both the verbal [$t(25) = 4.73$, $p < .001$] and the visual [$t(25) = 3.99$, $p < .01$] interference conditions. There was no significant difference in accuracy for cross-category pairs between the visual and the verbal interference conditions [$t(25) = 0.74$, $p > .05$]. For within-category pairs alone, accuracy was again significantly better in the no-interference condition than in both the visual [$t(25) = 4.55$, $p < .001$] and the verbal [$t(25) = 4.62$, $p < .001$] interference conditions. The difference in accuracy for within-category pairs between the verbal and the visual interference conditions was not significant [$t(25) = 0.01$, $p > .05$].

Experiment 2B: *Same-Different*

Figure 2, panel B shows the data from Experiment 2B. Essentially the same patterns of results and significance were found as those in Experiment 2A: No significant effect of verbal interference on CP was found. There was a small effect of visual interference, but this was mostly the result of a slightly better accuracy for within-category pairs than for cross-category pairs. The pair type $\times$ interference interaction did not approach significance [$F(1,25) = 0.39$, $MS_e = 0.0054$, $p = .681$].

Shift Toward Prototype

Figures 2C and 2D show the within-category accuracy scores broken down by the order in which the boundary

Figure 2. Panels A and B show accuracy in Experiments 2A and 2B for cross- and within-category pairs. Panel A shows Experiment 2A (two-alternative forced choice). Panel B shows Experiment 2B (*same-different*). Panels C and D show accuracy in Experiments 2A and 2B for within-category pairs when the boundary stimulus was the target and when it was the test. Panel C shows Experiment 2A. Panel D shows Experiment 2B. Error bars show ± 1 standard error in all panels.

and nonboundary stimuli occurred. It can be seen that there was clearly no order effect for any condition for the 2AFC task (Figure 2, panel C) but that there was for all the conditions for the *same-different* task (Figure 2, panel D). Boundary followed by nonboundary stimuli was more accurate than the reverse [$F(1,25) = 48.62$, $MS_e = 0.0018$, $p < .001$]. The effect of interference was also significant [$F(2,50) = 4.35$, $MS_e = 0.001$, $p < .05$].

Direct Language Model

If one compares the predictions of the simple direct language model in Table 1 with Figures 2A and 2B, it can be seen that the predictions are very close to the observed scores for the cross-category condition for both 2AFC (observed = 79.17%; predicted = 78%) and *same-different* judgments (observed $A' = .76$; predicted $A' = .79$). Predictions for within-category trials were much too low.

To summarize, for both tasks, there was a significant advantage in the discrimination of cross-category pairs over within-category pairs in the no-interference condition. The advantage persisted in both interference conditions, although overall accuracy was lower in these conditions than for baseline. Neither interference condition had a significant effect on CP. No support was found for the

shift-toward-prototype account of CP, but the same order effect for within-category *same-different* trials was found as in the previous experiment. The simple direct language model predicted cross-category performance well but underestimated within-category performance.

Discussion

Unlike Experiment 1, these experiments did not reproduce the effect reported in Roberson and Davidoff (2000). With the order of interference conditions blocked (Experiments 1A and 1B), verbal interference eliminated CP, whereas visual interference reduced accuracy overall but did not eliminate CP. With the order of conditions randomly interleaved so that interference was no longer predictable (Experiments 2A and 2B), there was no effect of verbal interference on CP.

This finding suggests that the account given in Roberson and Davidoff (2000) is incomplete. Color CP can survive verbal interference as long as the interference cannot be predicted. Rather than disrupting the retention of verbal color codes, predictable verbal interference may determine whether participants use a labeling strategy. This is consistent with previous findings that the use of labeling strategies in color judgments can be manipulated by varying their likely benefits (e.g., Bornstein,

1976; Kay & Kempton, 1984). For instance, Kay and Kempton (Experiment 2) found that a verbal bias effect in a color difference judgment task was removed by altering the instructions and the stimulus viewing conditions to make a verbal strategy less viable.

The cross-category data for both tasks were again close to the predictions from the direct-naming model, but the within-category predictions were again an underestimate. This is consistent with within-category choices' being based mostly on comparisons of perceptual representations. On the other hand, cross-category decisions could be based entirely on comparison of verbal labels or on some combination of verbal labels and visual representations. However, given that the sizes of the visual and the verbal interference effects are about the same and that they affect within- and cross-category decisions about equally, both kinds of comparisons may be involved in both kinds of decisions.

The foregoing explanation—CP elimination due to strategy change—is consistent with a direct language account and with the findings from Experiments 1 and 2. However, it is predicated on the assumption that verbal interference has no effect on color name retention. Experiment 3 tested this assumption.

EXPERIMENT 3

Interference and Name Retention (3A, Blocked; 3B, Randomized)

Roberson and Davidoff (2000) assumed that verbal interference during the ISI leads to a loss in retention of the verbal code, whereas visual interference does not. Such retention loss could be due to competition for use of the articulatory loop between name retention and performing the verbal interference task (e.g., Baddeley, 1986; Baddeley & Hitch, 1974). Verbal interference does make retention of verbal material harder when several items have to be remembered (e.g., Ternes & Yuille, 1972). However, it is less clear that retention of a single word would be similarly affected. Experiment 3 tested this possibility by examining the resilience of verbal color codes to verbal and visual interference. A 2AFC delayed color-to-name matching task was used that required that the name of a target color be selected after the ISI, during which no interference, verbal interference, or visual interference was included. If verbal interference hinders retention of a name code, as Roberson and Davidoff's account assumed, accuracy should be reduced within that condition, relative to the no-interference and visual interference conditions.

The experiment was run in two versions mirroring the difference between Experiments 1 and 2. In Experiment 3A, interference type was blocked, whereas in Experiment 3B, interference type was randomized within blocks. If verbal interference were to hinder name retention, relative to the other types of interference in the blocked design, the unblocked design would provide a check that this did not result from attempts to circumvent the effects of verbal interference.

Method

Participants

Twenty-two different participants (3 men and 19 women; age range, 18–32 years) took part in each experiment. All were students at the University of Surrey. Some obtained course credit for participation. None had taken part in the previous experiments.

Stimuli and Apparatus

Four color stimuli were used, two examples each of *blue* and *green*. For each category, one stimulus was a good example of the category (focal stimuli), and the other was a less good example taken from closer to the blue–green boundary (boundary stimuli). The focal stimuli were the most saturated examples of those colors that could be displayed on the monitor. The boundary stimuli were drawn from the stimuli used in the previous experiments: blue, 8.76 BG 6.20/7.40; green, 6.06 BG 6.20/7.40. For the verbal and visual interference tasks, a subset of the word and shape pairs used in Experiments 1 and 2 was used.

Procedure

On each trial, the target color was shown for 1,000 msec, followed by a 5,000-msec filled or unfilled ISI. During the ISI, there was no interference, verbal interference, or visual interference. The target was chosen at random for each trial, but each stimulus was shown equally often in each condition. Two test stimuli were then presented side by side for 1,000 msec. Each of the test stimuli consisted of a square black border with either BLUE or GREEN printed inside in black uppercase Arial font (size 38) on a neutral background. The position of the target name was chosen randomly for each trial, but it appeared on the left and the right equally often. The name of the target had to be chosen, and the choice indicated using the left or the right mouse button, as appropriate. For Experiment 3A, there were six blocks of 16 trials—two blocks of each of the three conditions in random order. In Experiment 3B, interference was randomized within each block. The interference tasks were the same as those for the two previous experiments, but only subsets of the stimuli were used. Ten practice trials were performed before each condition. Auditory feedback was given to indicate errors on practice trials; no feedback was given on experimental trials.

Results

On the interference tasks, the means across participant accuracy and reaction times were 77.70% and 991 msec and 85.23% and 897 msec for visual and verbal interference, respectively.

Figure 3. Experiment 3: Accuracy for focal and boundary stimuli for each interference condition. Error bars show ± 1 standard error.

The mean percentages of correct responses for the primary task for focal and boundary colors are shown in Figure 3. The results were analyzed with a repeated measures two-way ANOVA, with focality (focal or boundary) and interference (none, verbal, or visual) as factors. Focal colors were named more accurately than boundary colors [$F(1,21) = 99.11$, $MS_e = 212.86$, $p < .0001$]. There was no significant effect of interference type [$F(2,42) = 0.90$, $MS_e = 56.70$, $p > .05$] and no focality $\times$ interference interaction [$F(2,42) = 1.53$, $MS_e = 71.95$, $p > .05$]. Post hoc (protected t) tests showed that accuracy for focal colors was significantly greater than for boundary colors in all the interference conditions. There were no significant differences between baseline performance and either interference condition for either focal or boundary colors.

The results from Experiment 3B showed the same pattern of scores and significant differences.

Discussion

Accuracy was greater for focal stimuli than for boundary stimuli under all the interference conditions. However, since naming is more reliable for nonboundary than for boundary stimuli (see Experiment 1), accuracy differences probably reflect differences in the reliability of naming focal and boundary stimuli, rather than differences in the resilience of the name codes.

Interference had no effect on the accuracy of naming. This result challenges Roberson and Davidoff's (2000) account of the effect of verbal interference on CP. Their account presumes that the verbal interference task and retention of the color name compete for common processing resources, such as the phonological loop, and that the two tasks cannot be accomplished successfully at the same time. The present results suggest that the two tasks can be accomplished at the same time, under some circumstances at least. The question then arises as to why verbal interference eliminated CP in Experiments 1A and 1B and in Roberson and Davidoff. The present result reinforces the strategic bias suggestion made earlier (see the Discussion section, Experiments 2A and 2B). There, we suggested that when interference type was known in advance, the participants chose not to use a naming strategy. Perhaps coping with verbal interference made name retention more effortful, and an alternative route of coping with the primary task was used. In the present task, because the response required name recognition, no simple alternative route was available.

GENERAL DISCUSSION

The results from Experiments 1A and 1B essentially replicated Roberson and Davidoff (2000). Consistent with their direct language account of CP, visual interference did not affect CP, but verbal interference did. The methods used in the present experiments differed from Roberson and Davidoff's method in a number of ways. Their visual interference task required no overt response, whereas their verbal interference task did. This may have

led participants to regard the visual task as less important than the verbal interference task. We cannot tell, because performance levels were not recorded on either interference task. The present interference tasks were based on the theory of working memory (Baddeley & Hitch, 1974) and required the same overt response, and performance levels were recorded to emphasize the equal importance of the two interference tasks. The secondary task data were used in a covariate analysis with the primary task data in order to check that performance on the primary task was not affected by varying effort on the secondary task. The present experiments also used the *same-different* and the 2AFC methods as a further test of the generality of the findings. Roberson and Davidoff's main findings were replicated despite the foregoing variations on their methods and data analysis, testifying to the robustness of the effects and supporting their direct language account of CP.

The direct-naming account was further supported by the tests of the simple direct-naming model. Its predictions matched the experimental data well for the no-interference condition for cross-category trials, but not for within-category trials. The model had been intended only to provide an upper bound for accuracy based on name comparisons alone. If observed performance had been more accurate than predicted performance, this may have indicated the degree to which visual comparisons were also used on cross-category trials. (The visual comparison could be of perceptual representations, of their representations in memory, or of a perceptual representation with a memory representation; henceforth, however, we will use *visual representation* as shorthand). Despite the attractions of parsimony, it seems likely that visual information would also influence decisions when name comparisons do not produce a clear choice. This will happen often on within-category trials, in which two stimuli will be given the same name (both test stimuli for 2AFC trials or the target and the test stimuli for *same-different* trials). It will also happen on cross-category trials in which one of the two stimuli is named incorrectly. The model assumes that on these equivocal trials, the choice will be random. However, it seems more likely that visual comparisons would be used to resolve the uncertainty. The naming data we used to predict the upper bound may have overestimated naming reliability in the experimental trials. In the naming study, only a single color had to be named, whereas in the experimental trials, up to three stimuli had to be named. This may have reduced naming reliability, because it required more processing resources or because the extra stimuli acted as reference points and affected naming. If the foregoing is true, there is scope in the model for providing a role for visual comparisons as an adjudicator on equivocal trials.

However, the results of Experiments 2A and 2B suggest that the simple direct language account needs qualifying. In Experiments 1A and 1B, the type of interference was known on each trial. In contrast, in Experiments 2A and 2B, the three interference conditions were randomly inter-

leaved, producing uncertainty about the type of interference on a given trial; yet CP occurred in all the interference conditions. The difference between the results of the two pairs of experiments suggests that foreknowledge of the type of interference affects how the primary task is performed. If verbal interference is expected, visual codes are used, whereas if visual interference is expected, verbal codes are used. Where no interference is expected, both codes may be used. This strategic switching could be accomplished by naming the test stimulus on visual interference trials but not naming it on verbal interference trials. Alternatively, it may reflect whether name rehearsal occurs during the ISI. The former seems more probable than the latter, because even when the type of interference was not known at the beginning of a trial, it would have been known when the interference task started. Thus, if verbal rehearsal was not possible with verbal interference or was believed to be impossible, it would have stopped at the onset of verbal interference in Experiments 2A and 2B. Given that CP occurred with verbal interference in these experiments, naming the test stimulus without subsequent rehearsal may be sufficient to produce CP.

However, in view of the foregoing argument, the direct-naming account requires that the memory representation of the name can survive relatively intact across the ISI, despite verbal interference. This was tested in Experiments 3A and 3B. Test stimuli had to be matched to verbal labels after a 5-sec ISI that contained either no interference, visual interference, or verbal interference. Interference type was blocked in Experiment 3A and randomized in Experiment 3B. Interference did not reduce accuracy in any condition in either experiment. These results, then, raise the question of why the target was not named in Experiments 1A and 1B when it was known that verbal interference would occur. One possibility is that whereas verbal interference during the ISI need not affect accuracy of color name retention, the participants believed that it did, and they therefore chose not to label the test stimulus. Alternatively, retaining the name during verbal interference may be effortful, and the participants chose an easier route.

We had predicted from Huttenlocher et al.'s (2000) theory that there should be an order effect for within-category trials. This was based on the assumption that the effective category prototype was the preexisting category prototype and that this lay further from the category boundary than did any of the experimental stimuli. If the memory representation of the target stimulus shifts toward the prototype during the ISI, this should reduce its separation from the representation of the test stimulus when the boundary stimulus was the target but should increase the separation when the nonboundary stimulus was the target. Therefore, the latter case should be more accurate than the former. Exactly the opposite was found. One possible explanation is that the order effect is due to a shift toward the prototype but the effective prototype is on the category boundary. We assumed that

the relevant prototypes were those of the preexisting blue and green categories—the best exemplars—and, therefore, were further out along our color axes (away from the boundary) than any of the experimental stimuli. In doing so, we may have extended Huttenlocher et al.'s model inappropriately. They considered the case of temporary categories, such as those that might be formed of a set of screws, varying in size, seen for the first time. Some overall representation of the set might be formed, including measures of the average and of the dispersion. Now, in remembering a particular instance, judgment may be biased toward the set average or temporary category prototype (see also Spencer & Hund, 2002). The equivalent in our case would be the average of the stimuli used in the *same-different* tasks, and this centered on the category boundary. If the memory shift is toward the temporary category average, this would account for the advantage of presenting the boundary stimulus first. The argument is exactly the same as that for the predicted order effect, except that now the direction of shift is toward the boundary, rather than away from it. However, prototype bias now cannot account for CP. For cross-category pairs, whichever is the target, the shift toward the prototype takes it in the same direction as the test stimulus, which should reduce accuracy, rather than increase it, as required for CP. Thus, prototype bias seems to be operating, but it cannot be the cause of CP.

Perhaps more puzzling is that there was no evidence of an order effect and, hence, prototype bias in the 2AFC task. One reason for including 2AFC and *same-different* tasks was that, although they are often regarded as equivalent, the former may invite a discriminative response, whereas the latter may invite a categorical response. These rather vague ideas have a long history in the context of the categorical perception of speech. Pisoni (1973) referred to them as auditory mode and phonetic mode, and Guenther, Husain, Cohen, and Shinn-Cunningham (1999), following Durlach and Braida (1969), used the terms sensory-trace mode and context-coding mode. Perhaps there is more scope for the influence of the prototype in categorical mode than in discriminative mode. On the other hand, perhaps it is just that repetition of the target as one of the test stimuli in the 2AFC task reactivates or adds to the target's fading trace. Huttenlocher et al. (2000) proposed that the influence of the prototype would grow as the memory representation of the target fades. Perhaps the reactivation of the trace leaves little scope for prototype influence. On *different* trials for the *same-different* task, the target is not presented at test, and no such reactivation of its trace can occur.

Finally, although we believe that our data show that categorical perception in delayed discrimination tasks may be due to comparison of verbal labels, rather than warping of perceptual color space, this does not mean that true CP does not occur under other circumstances. More conservatively, our data and Roberson and Davidoff's (2000) may just show that unless the target is named, CP does not occur. This may be because labels

are then not available to be used in a comparison process, but it could also be because labeling the target activates a category code that is then available to be used as part of the choice process. This would be consistent with Bornstein and Korda's (1984) view that the target activates physical and categorical codes in parallel and with Pisoni's (1973) and Guenther et al.'s (1999) views on CP for speech and nonspeech sounds, respectively.

REFERENCES

- AHISSAR, M., & HOCHSTEIN, S. (1997). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences*, **90**, 5718-5722.
- BADDELEY, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-87). London: Academic Press.
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BORNSTEIN, M. H. (1974). Perceptual generalization: A note on the peak shift. *Psychological Bulletin*, **81**, 804-808.
- BORNSTEIN, M. H. (1976). Name codes and color memory. *American Journal of Psychology*, **89**, 269-279.
- 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., KESSEN, W., & WEISKOPF, S. (1976). Color vision and hue categorization in young infants. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 115-129.
- 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.
- BORNSTEIN, M. H., & MONROE, M. D. (1980). Chromatic information processing: Rate depends on stimulus location in the category and psychological complexity. *Psychological Research*, **42**, 213-225.
- BOYNTON, R. M., FARGO, L., OLSON, C. X., & SMALLMAN, H. S. (1989). Category effects in color memory. *Color Research & Application*, **12**, 107-123.
- BROWN, G. (1987). Phonological coding in rhyming and homophony judgement. *Acta Psychologica*, **65**, 247-262.
- CALDER, A. J., YOUNG, A. W., PERRETT, D. I., ETCOFF, N. L., & ROWLAND, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, **3**, 81-117.
- CAMPBELL, R., PASCALIS, O., COLEMAN, S. B., WALLACE, B., & BENSON, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proceedings of the Royal Society of London: Series B*, **264**, 1429-1434.
- CAMPBELL, R., WOLL, B., BENSON, P., & WALLACE, S. B. (1999). Categorical perception of face actions: Their role in sign language and in communicative facial displays. *Quarterly Journal of Experimental Psychology*, **52**, 67-95.
- DAVIES, I. R. L., & FRANKLIN, A. (2002). Categorical similarity may affect colour pop-out in infants after all. *British Journal of Developmental Psychology*, **20**, 185-203.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.
- FAHLE, M. (1994). Human pattern recognition: Parallel processing and perceptual learning. *Perception*, **23**, 411-427.
- FAHLE, M., EDELMAN, S., & POGGIO, T. (1995). Fast perceptual learning and hyperactivity. *Vision Research*, **35**, 3003-3013.
- FLETCHER, R. (1980). *The City University Color Vision Test* (2nd ed.). London: Keeler.
- FRANKLIN, A., & DAVIES, I. R. L. (2002). *Categorical colour perception in four-month infants*. Manuscript submitted for publication.
- FUJISAKA, H., & KAWASHIMA, T. (1971). On the modes and hearing mechanisms of speech perception. In *Annual report of the Engineering Research Institute* (Vol. 28, pp. 67-73). University of Tokyo, Faculty of Engineering.
- GERHARDSTEIN, P., RENNER, P., & ROVEE-COLLIER, C. (1999). The roles of perceptual and categorical similarity in color pop-out in infants. *British Journal of Developmental Psychology*, **17**, 403-420.
- GOLDSTONE, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, **123**, 178-200.
- GOLDSTONE, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, **6**, 298-304.
- GOLDSTONE, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, **49**, 585-612.
- GOLDSTONE, R. L., LIPPA, Y., & SHIFFRIN, R. M. (2001). Altering object representations through category learning. *Cognition*, **78**, 27-43.
- GUENTHER, F. G., HUSAIN, F. T., COHEN, M. A., & SHINN-CUNNINGHAM, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, **106**, 2900-2912.
- 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). Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Kaiserslautern, FRG.
- HUNT, R. W. G. (1987). *Measuring color*. Chichester, U.K.: Ellis Horwood.
- HUTTENLOCHER, J., HEDGES, L. V., & VEVEA, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, **129**, 220-241.
- KAY, P., & KEMPTON, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, **86**, 65-78.
- 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.
- LIVINGSTON, K. R., ANDREWS, J. K., & HARNAD, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 732-753.
- MACMILLAN, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 53-85). Cambridge: Cambridge University Press.
- MALT, B. C., SLOMAN, S. A., GENNARI, S., SHI, M., & WANG, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory & Language*, **40**, 230-262.
- MUNSELL, A. H. (1923). *A color notation*. Baltimore: Munsell Color Co.
- 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.
- 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 non-speech 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.
- RACK, J. P. (1985). Orthographic and phonetic coding in developmental dyslexia. *British Journal of Psychology*, **76**, 325-340.
- 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. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, **93**, 10-20.

- ROSEN, S., & HOWELL, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 113-160). Cambridge: Cambridge University Press.
- SANDELL, J. H., GROSS, C. G., & BORNSTEIN, M. H. (1983). Color categorization in macaques. *Journal of Comparative & Physiological Psychology*, **93**, 626-635.
- SCHYNS, P. G., & RODET, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 681-696.
- 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.
- SPENCER, J. P., & HUND, A. M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General*, **131**, 16-37.
- SWETS, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics*. Mahwah, NJ: Erlbaum.
- TAJFEL, H., & WILKES, H. (1963). Classification and quantitative judgment. *British Journal of Psychology*, **54**, 101-114.
- TERNES, W., & YUILLE, J. (1972). Words and pictures in an STM task. *Journal of Experimental Psychology*, **96**, 78-86.
- TRAVIS, D. (1991). *Effective color displays: Theory and practice*. London: Academic Press.
- UCHIKAWA, K., & SHINODA, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research & Application*, **21**, 430-439.
- WALSH, V., KULIKOWSKI, J. J., BUTLER, S. R., & CARDEN, D. (1989). The effects of lesions of area V4 on the visual abilities of macaques: Colour categorization. *Behavioural Brain Research*, **52**, 81-89.
- WYSZECKI, G., & STILES, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.
- 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.
- ZATORRE, R. J., & HALPERN, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, **26**, 384-395.

(Manuscript received May 1, 2002;
revision accepted for publication March 4, 2003.)