

# Color categories only affect post-perceptual processes when same- and different-category colors are equally discriminable

Xun He,<sup>1,2,4</sup> Christoph Witzel,<sup>2</sup> Lewis Forder,<sup>2</sup> Alexandra Clifford,<sup>3</sup> and Anna Franklin<sup>2,5</sup>

<sup>1</sup>Psychology Research Centre, School of Design, Engineering and Computing, Bournemouth University, BH12 5BB, UK

<sup>2</sup>The Sussex Colour Group, School of Psychology, University of Sussex, BN1 9RH, UK

<sup>3</sup>Department of Psychology, University of Surrey, Guildford, Surrey GU2 7XH, UK

<sup>4</sup>e-mail: xhe@bournemouth.ac.uk

<sup>5</sup>e-mail: anna.franklin@sussex.ac.uk

Received September 30, 2013; revised December 13, 2013; accepted January 6, 2014;  
posted January 17, 2014 (Doc. ID 198626); published 0 MONTH 0000

Prior claims that color categories affect color perception are confounded by inequalities in the color space used to equate same- and different-category colors. Here, we equate same- and different-category colors in the number of just-noticeable differences, and measure event-related potentials (ERPs) to these colors on a visual oddball task to establish if color categories affect perceptual or post-perceptual stages of processing. Category effects were found from 200 ms after color presentation, only in ERP components that reflect post-perceptual processes (e.g., N2, P3). The findings suggest that color categories affect post-perceptual processing, but do not affect the perceptual representation of color. © 2014 Optical Society of America

OCIS codes: (330.1720) Color vision; (330.5020) Perception psychology; (330.5510) Psychophysics.

<http://dx.doi.org/10.1364/JOSAA.31.00000A>

## 1. INTRODUCTION

Although we are able to perceive more than two million different colors [1,2], colors are commonly grouped into a number of more or less discrete categories (e.g., red, green, blue). The effect of categorization on how color is perceived has been extensively researched [3,4]. Numerous studies have claimed that the phenomenon of “categorical perception,” where discrimination of stimuli from different categories is easier than discrimination of an equivalent same category stimulus difference [5], extends to the domain of color. For example, some have suggested that people are faster or more accurate at searching for a colored target when target and distractors are from different color categories than when they are from the same category, even when same- and different-category chromatic differences are equated (e.g., [6–8]). Other potential categorical effects have been documented when people memorize colors or judge their similarity (e.g., [9,10]).

There has been uncertainty about whether color category effects really do occur at a perceptual level, as the term “categorical perception” would suggest [11–15]. Psychophysical studies have disagreed whether categorical effects are present at detection threshold [13] or not [14]. Distinguishing different-category colors may be easier than same-category colors because the categorical difference enhances attention [15]. Alternatively, a same-category disadvantage could arise because same-category colors have the same name despite being different colors, which leads to conflict at the stage of decision making [11]. Color categories may therefore affect post-perceptual stages of processing, but the underlying perceptual representation of color and early stages of color processing (e.g., at visual cortex) may remain unaffected.

In order to clarify what stages of color processing are affected by color categories, a number of studies have employed the event-related potential (ERP) technique [15–24]. This technique uses electrodes to measure electrical activity from the scalp elicited in response to sensory, cognitive, or motor events. The resulting ERPs are then represented as waveforms through time at various locations on the scalp surface. Typical visual ERPs (i.e., ERPs generated by visual stimuli) have a series of components (P1, N1, P2, N2, P3), which are termed after their polarity (P = positive, N = negative) and relative positions in the waveforms (e.g., P1 means the first positive component) and which are known to index certain sensory or post-perceptual processes [25,26].

Studies that have used the ERP technique to investigate color category effects have recorded ERPs while participants search for colors [18], while they passively or actively detect changes in the color presented [16–23], or while they make judgments about whether colors are the same or different [24]. The majority of these studies have claimed that the influence of categories can be seen at early perceptual stages of the time course, during the first couple of hundred milliseconds after color presentation [16–23]. For example, in one study, the detection of a color change elicited an ERP component around 100 ms (P1) that peaked roughly 5 ms earlier when the color change was different- than same-category [17]. Others have claimed that category effects are pre-attentive or unconscious and are found for an ERP component (the visual mismatch negativity, vMMN) [27,28] that is elicited even when attention is not directed to the colors [19–21]. An fMRI investigation of categorical effects in visual search that found greater activation in V2/V3 for different- than same-category

color search has strengthened such claims [29]. Such findings have been taken to suggest that color categories affect perceptual processing at visual cortex [16,17,19].

However, there is also concern over whether these so-called category effects are actually categorical at all because these studies only coarsely controlled color differences using the Munsell color system, or CIELUV/CIELAB color spaces. There are known perceptual inequalities in these color spaces [30,31], such that color differences equated in Munsell hue or Euclidean distance in CIELUV/LAB are not necessarily equally discriminable. It is therefore possible that the early perceptual category effects in some ERP studies may simply be explained by failure to control the discriminability of the colors [32].

Nevertheless, there is also evidence for color category effects that cannot easily be explained by perceptual inequalities in the color spaces used to equate same- and different-category colors or by other stimulus issues. First, effects are likely to be categorical when there are cross-linguistic differences in the effect. For example, if there is an effect only for speakers who have separate terms for the categories, then there is greater certainty that the effect is a result of categories, as if the effect were due to problems with the color spaces used to equate colors then the effect would be present for all participants irrespective of the terms in their language [19,33–35]. One ERP study examined the time course of the effect of cross-linguistic differences in color terms (Greek and English) and found cross-linguistic differences within the time range of an early perceptual and pre-attentive ERP component (the vMMN) [19]. However, others have also questioned whether these effects really were independent of attention or whether, at least for some participants (particularly the English speakers), the vMMN ERP component was actually obscured by an attentional one (the “N2b”) [15–17,20]. This means that there is still some uncertainty about whether cross-linguistic differences in color category effects are due to differences in perception or post-perceptual processes such as attention.

Second, effects are likely to be categorical when they arise following the learning of novel categories (e.g., if participants are trained to divide greens into two new categories). In such studies, the stimuli are the same before and after participants learn the novel categories, and so effects induced by category learning cannot be due to stimulus issues [15,21,36]. One ERP study examined the time course of category effects induced following the learning of novel categories [15] and only found evidence for category effects in post-perceptual stages of processing (P3 component, 350–600 ms). However, it does remain possible that category effects would appear earlier in perceptual processing if the newly learnt color categories were more familiar.

A third method of ensuring that “category” effects cannot be accounted for by inequalities in the color space used to equate same- and different-category colors is to instead equate colors in empirically measured just-noticeable differences (JNDs). A JND is the minimal difference between two colors that an observer is just able to perceive, and equating the number of JNDs for same- and different-category color differences controls for discriminability. A set of studies that have previously taken this approach have found color category effects [37–39]. For example, performance in a speeded discrimination task increased toward category boundaries when colors from same- and different-categories were equally

discriminable [37–39]. One potential explanation for these category effects is that when equally discriminable colors are distinguished on speeded tasks such as search, memory, or change detection tasks, categorical distinctions affect post-perceptual processes (e.g., attention), which modulate performance. An alternative account is that the category effects result from categories modulating perception in a top-down manner. As colors are equated in discriminability, it might seem circular to argue that categories could influence perception. However, even though colors are equated in discriminable differences at threshold, it could still be possible for categories to modulate perception when participants make judgments about supra-threshold color differences under speeded conditions such as visual search or change detection.

The aim of the current study is to investigate the time course of color category effects in ERPs, when the discriminability of colors is carefully controlled through the measurement of discrimination thresholds. We investigate category effects across the blue–green category boundary, as this boundary has been the main focus of prior claims for color categories affecting perception, e.g., [6,7,9,10,16–18]. The current study has two stages: a stimulus definition stage and a main experiment. The stimulus definition stage aimed to define a stimulus set with blue/green color pairs that are equated in JNDs for the average participant. To get an idea of the average location of the blue–green category boundary and its range across participants we asked one group of participants to name colors in the blue–green region. Discrimination thresholds in terms of JNDs were then measured for a series of blue and green colors with another group of participants. Based on these measurements, we created equally discriminable color pairs that were expected to vary in categorical relationship (same-/different-category) for individuals in the main experiment. Separate groups of participants completed the stimulus definition naming and JND tasks, and participants were also different to those who completed the main experiment to ensure that there were no carry-over effects between tasks.

The main experiment aimed to establish the time course of category effects for the equally discriminable colors. To assess the category effect, a third group of participants viewed the colors on a visual oddball task [17]. The visual oddball task required participants to detect infrequently presented colors (the deviant stimuli) amongst frequent presentations of another color (the standard stimulus). Participants completed the visual oddball task twice: First when ERPs were recorded and participants counted the number of deviant stimuli. Participants did not make a manual response, which removed the possibility of having undesirable contamination from movement-related brain activities on the EEG data. Participants then completed the visual oddball task again when participants manually responded to the stimuli and we recorded reaction time (RT) and accuracy but not ERPs. Following this, the same participants named the colors used on the visual oddball task. The standard-deviant color pairs were classified as same- or different-category for each individual on the basis of their own color naming. This ensured that individual variation in the location of the blue–green category boundary was accounted for.

A number of ERP components are elicited on a visual oddball task. The early ERP components, namely P1 and N1, which typically occur 100–200 ms post-stimulus, generally

reflect the processing of visual information at a sensory level; thus they are called sensory-level components because they are sensitive to physical properties of visual stimuli [40–42]. Later activities, which include P2, N2, and P3, involve post-perceptual processes such as feature evaluation and context updating [43–45]. Studies have shown that these late components are enhanced by low-frequency stimuli [45–47]. An effect of color categories on the viewing of colors on the visual oddball task is characterized by differences in the peak latency or amplitude of the ERP components elicited by deviant colors that were from the same- versus different-category to the standard. The ERP components for which such category effects are found indicate the time course and nature of the effect of color categories. If categories affect post-perceptual processes, category effects should be found for post-perceptual ERP components (e.g., P2, N2, P3, components after 200 ms). If color categories affect perceptual processes, category effects should be found for perceptual ERP components (e.g., P1, N1).

## 2. STIMULUS DEFINITION

### A. Participants

Ten native British-English speakers took part in a color naming task [6 females, mean age = 23.5, standard deviation (SD) = 6.2]. Data from 17 participants (12 females, mean age = 23.2, SD = 5.0) were used to determine the JND values. All measurements from two further participants were removed because their JNDs were more than 3 SDs away from the group mean for at least one tested color. This removal of outlier data ensures a more accurate estimate of JNDs. All participants had normal or corrected-to-normal vision and had no color vision deficiencies on the Ishihara Color Vision Test [48] or the City Color Vision Test [49] (as for participants in the main experiment). All parts of the study were approved by the Life Sciences and Psychology Cluster based Ethics Committee at the University of Sussex (ref: AF0811) and the European Research Council Executive Agency Ethics Review Board (ref: 283605).

### B. Apparatus and Set-Up

MATLAB (MathWorks, Inc.) and Psychophysics Toolbox 3 (see [50,51]) were used to prepare the color naming and JND measurements. Stimuli were displayed on a 22" CRT monitor (color resolution: 8 bits/channel, spatial resolution: 1600 × 1200, refresh rate: 100 Hz). The CIE1931 chromaticity coordinates and luminance of the monitor primaries were:  $R = (0.626, 0.337, 14.24)$ ,  $G = (0.281, 0.614, 45.51)$ , and  $B = (0.151, 0.071, 5.28)$ . Gamma corrections were applied based on the measured gamma curves of the monitor primaries.

### C. Color Naming and Threshold Discrimination Experiments

Color naming and threshold discrimination experiments were run in order to define the colors for the main visual oddball experiment. The aim of the color naming experiment was to establish the average location of the blue–green boundary and its range across participants. This allows us to define a stimulus set, which is likely to provide color pairs that participants who take part in the main experiment will consider to be same- or different-category. A group of participants named 120 colors on an isoluminant circle ( $3^\circ$  steps) in the CIELUV

space at a lightness level typical for green ( $L^* = 50$ ) [52] on a gray background ( $L^* = 70$ ) with an equal-energy white point. Participants determined the name of the presented color by pressing one of eight keys (for the eight basic color names). The average boundary between green and blue was quantified as the hue angle for which participants named the two adjacent colors with equal probabilities (linear interpolation was performed when the actual data showed ambiguous boundaries). This average boundary was determined in the current setup at the hue azimuth of  $183^\circ$  (standard error =  $3^\circ$ ).

Discrimination thresholds were then measured in the blue–green region of color space around the average category boundary so that color pairs could be defined that were equated in the number of JNDs. For the JND measurements, 12 colors were tested between  $155^\circ$  and  $210^\circ$  hue angles at  $5^\circ$  steps in the CIELUV space because this corresponds approximately to an average JND in CIELUV according to previous data for slightly different colors [53]. The threshold discrimination task was the same as in Witzel and Gegenfurtner [53]. It was a color discrimination task with a 4-alternative forced choice response (4AFC), which required participants to identify the one stimulus out of four that was different to the rest. Stimuli were presented as discs  $1.9^\circ$  in diameter positioned  $3.5^\circ$  away from fixation (center-to-center) in a 2 by 2 grid layout. The task was combined with a 3-up-1-down staircase that converged toward the JNDs. Two staircases were measured for each test color, one toward blue and the other toward green. For each staircase we measured seven reversal points as this has provided a good compromise of stable JND estimates and length of experiment in prior research [53]. JNDs were calculated as the average of the data from the last six reversal points, excluding the first reversal point to avoid contamination from spurious responses. The JNDs from all the participants were then averaged and used to equate color pairs. The JND results are shown in Fig. 1.

Based on the JND measurements, four hues were defined (A, B, C, D; Table 1, see also Fig. 1) so that neighboring colors were separated by 3 JNDs. These hues were selected so that they spanned the average blue–green category boundary in order to maximize the possibility that participants in the main experiment would classify one of the color pairs as different-category.

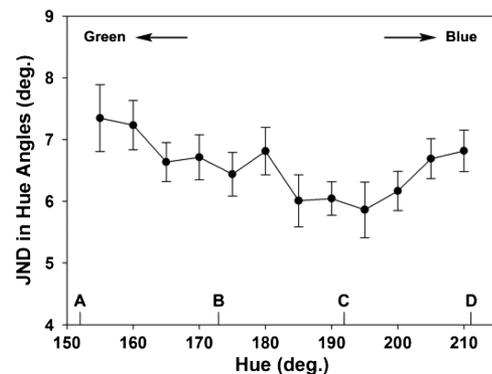


Fig. 1. JND results. Twelve test colors ( $155^\circ$ – $210^\circ$  in hue angle in the CIELUV space with  $5^\circ$  steps) were chosen around the average blue–green boundary ( $183^\circ$ ). For each test color, JND estimates were averaged to give the mean JND (in degree of hue angles). Error bars indicate standard errors (SEs) of JNDs. Based on the JND results, four colors (A, B, C, D), which are equated in number of JNDs, are defined for the main experiment.

**Table 1. Chromaticity Coordinates ( $x, y$ , CIE1931) for Stimuli and the White Point of the Monitor<sup>a</sup>**

Stimulus	$x$	$y$
A	0.272	0.433
B	0.237	0.380
C	0.222	0.333
D	0.220	0.292
Background	0.333	0.333

<sup>a</sup>Luminance for stimuli A–D was  $Y = 9.21 \text{ cd/m}^2 (\pm 0.1)$ ; background  $Y = 20.37 \text{ cd/m}^2$ , white point  $Y = 50 \text{ cd/m}^2$ .

### 3. MAIN EXPERIMENT: METHOD

#### A. Participants

Twenty native British-English speakers (18 female, mean age = 21.0, SD = 3.0) participated in the main experiment. Their eyesight and color vision were tested beforehand.

#### B. Apparatus and Set-Up

Materials were prepared with e-Prime 2 (Psychology Software Tools, Inc.), and participants were seated comfortably in a dark room 77 cm away from a 22" CRT monitor (color resolution: 8 bits/channel, spatial resolution:  $1024 \times 768$ , refresh rate: 75 Hz). The CIE1931 chromaticity coordinates and luminance of the monitor primaries were  $R = (0.627, 0.339, 16.83)$ ,  $G = (0.280, 0.614, 42.98)$ , and  $B = (0.150, 0.071, 5.55)$ . All RGB values were gamma-corrected based on the inverted gamma functions of the monitor.

#### C. Procedure

The main experiment had three parts: electroencephalogram (EEG) and behavioral versions of the visual oddball task and a color naming task. First, participants completed the EEG version of the visual oddball task. The visual oddball task consisted of eight 100-trial blocks. On each trial, a single color square (color A, B, C, or D;  $7.5^\circ \times 7.5^\circ$ ) was presented for 400 ms in the screen center. The square was presented on the same gray background as used in the stimulus definition experiments, followed by a random interval of 1200–1600 ms during which only the gray background was presented. The square could be of one of three colors: the standard (a frequently presented color) and two deviants (infrequently presented colors). The standard color was presented eight times at the beginning of each block to ensure that it was familiar to the participants and that they could identify it as the standard. For the rest of the block, the standard and two deviant colors were presented in a pseudo-random order so that there were no consecutive deviant colors. The number of the deviant trials in a block could be 15, 16, 17, or 18 with equal possibilities across the eight blocks. The four colors (A, B, C, D) formed two stimulus sets, each of which consisted of a standard color (the central color in the set) and two deviants (the outer colors in the set, Fig. 2). Half of the participants were randomly allocated to color set 1 and the other half to color set 2.

Before the experimental trials of the visual oddball task commenced, participants were familiarized with the procedure with a 100-trial practice block, in which reddish colors were presented on the gray background. Participants completed the visual oddball task twice. The first time, EEG was recorded and the participants' task was to count the

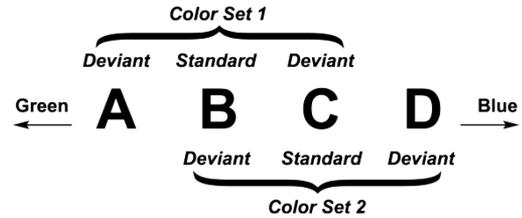


Fig. 2. Color sets used in the visual oddball experiment. There were four colors (A–D). Adjacent colors are equated in their discriminability. Two color sets were formed from these four colors (ABC or BCD). The color in the middle of each set was the frequent “standard” color, and the two outer colors in a set were the infrequent “deviant” colors. Deviant stimuli were categorized as same- or different-category to the standard stimulus on the basis of each individual’s color naming.

occurrences of the deviant colors and to orally report the number to the experimenter immediately after each block. An oral response was chosen over a manual one so that the EEG data would not be contaminated by movement artifacts or related brain activities.

The second time participants completed the visual oddball task EEG was not recorded and participants were asked to press one of two keys to indicate whether the stimulus was standard or deviant and RTs were recorded. Half of the participants pressed the “c” key for the standard color and “m” for the deviant colors, with the key allocation reversed for the other half of participants.

Since color-category boundaries vary across observers (as seen in the stimulus definition naming experiment), we tested each participant’s color naming of the standard and deviant colors that they viewed in the oddball task to accommodate individual differences. This ensures that standard and deviant color pairs are accurately classified as same- or different-category for each participant. Each color was presented as a square patch of the same size on the same background as in the oddball task. Each color remained on the screen until a response was made, followed by an intertrial interval of 1500 ms. Participants responded by pressing “c” and “m” for *green* and *blue*, with key allocation counterbalanced across subjects. In later analysis, each individual’s naming was then used to classify the deviant colors as same- or different-category to the standard for each participant separately.

#### D. Data Recording and Processing

EEG was recorded and processed with NeuroScan SynAmps<sup>2</sup> amplifiers and SCAN 4.3 software (NeuroScan/Compumedics, Inc.) at a digitizing rate of 500 Hz, and with a physical band-pass filter (0.1–100 Hz). Electrophysiological data were recorded from 62 electrodes that were held in place with an elastic cap. The locations of these 62 electrode sites on scalp surface followed the 10–5 nomenclature system (an extension to the 10–20 system) [54]. Additionally, eye blinks and eye movements were monitored with two bi-polar channels: two electrodes located at the outer canthi as the horizontal electrooculogram (EOG) channel and two electrodes above and below the left eye for recording vertical EOG activities. Apart from these two bi-polar channels, EEG responses at all scalp sites and the right mastoid were physically referenced to the left mastoid during data recording. Impedance of each channel was reduced below 5 k $\Omega$  before data collection.

After data collection, the recorded data were subjected to a zero phase-shift low-pass filter (amplitude cutoff frequency at

40 Hz, 48 dB/oct roll-off) and then digitally re-referenced to the mean voltage of electrical activities recorded at the two mastoids. Trials containing activities out of the range of  $\pm 60 \mu\text{V}$  were rejected as contaminated with artifacts. ERPs were then generated by averaging EEG activities over trials time-locked to stimulus onsets (for different experimental conditions separately). The averaged EEG segments lasted 900 ms, starting at 100 ms before stimulus onset and ending at 800 ms after stimulus onset and were measured relative to the 100-ms pre-stimulus baseline.

#### 4. MAIN EXPERIMENT: RESULTS

There were three stages of analysis. First, naming data were analyzed to establish the category membership of the deviant colors for individual participants. Second, behavioral data were analyzed with the main aim of establishing whether there was a category effect in participants' behavioral response. Third, EEG data were analyzed with the main aim of establishing whether there was a category effect for sensory or post-perceptual ERP components. The analysis is reported in this order (a reverse order to the reporting of the tasks in the methods, and the order of completion by participants) as this is the most logical order in which to present the findings.

##### A. Naming

One participant was ambiguous in the naming of the standard color (named green in 48% and blue in 52% of the trials). This participant's data were excluded from further analysis as their ambiguous naming of the standard meant that deviants could not be classified as same- or different-category for that participant. The blue–green category boundary was clear for the remaining 19 participants: for all colors the most frequent name given to a color was given on at least 60% of the trials. On the basis of the color naming in the stimulus definition phase of the study, it was expected that on average stimuli A and B would be named green and C and D would be named blue by many of the participants. This was indeed the most common location of the category boundary for participants in the main experiment as well (for 12 participants), yet there was also individual variation (seven participants, who saw colors B, C, and D in the oddball task, put the category boundary between C and D). Deviant colors were classified as same- or different-category to the standard color for each individual according to their own naming in order to take this individual variation in color naming into account.

##### B. Behavioral Oddball Task

The behavioral data (Section 4.B) and the EEG data (Section 4.C) from the visual oddball task were both analyzed using data from participants who saw either set 1 or set 2. Preliminary analyses with color set as a factor revealed no significant interactions of Set with Color,  $p = 0.49$  for hit rates and  $p = 0.98$  for RTs. Therefore the behavior analyses are presented for data collapsed across sets. For the behavioral oddball task data, trials with RTs more than three SDs away from the mean value were rejected for each participant. Hit rates and mean RTs for correct identification of the stimulus as standard or deviant for the three color conditions (standard, same-category deviant, different-category deviant) were then analyzed with one-way repeated-measures analyses of variance (ANOVAs). For hit rates, the main effect was significant,

$F(2, 36) = 30.97$ ,  $p < 0.0001$ . Least significance difference (LSD) pairwise comparisons revealed that the performance was better for the standard color (98.7%) than for the deviants (same-category 75.6%, different-category 80.4%),  $ps < 0.0001$ . The hit rates for the two deviant colors, however, did not differ,  $p = 0.14$ , showing no evidence of a category effect. This pattern was replicated in the RT data, where a significant main effect was found,  $F(2, 36) = 86.58$ ,  $p < 0.0001$ . RTs were significantly shorter for the standard color (406 ms) than for the deviant colors (same-category 515 ms, different-category 510 ms),  $ps < 0.0001$ . The RTs for the two deviant colors were comparable,  $p = 0.59$ , providing no evidence for a significant category effect.

##### C. EEG Oddball Task: Electrophysiology

Two participants elicited very strong alpha waves (8–13 Hz spontaneous EEG rhythmic activities) [55], which contaminated the ERP waveforms greatly. Additionally, another two persons showed an excessive number of artifacts in the EEG data, resulting in too few accepted trials for any ERP analysis. Data from these four subjects were removed from any ERP analysis. Therefore, 15 participants' data were analyzed and are presented here. ERP components and deflections were quantified as mean amplitudes at electrode sites within time windows that taken together showed the most prominent activities over space (electrode sites) and time (time windows) of the analyzed ERP responses. This was in line with the standard analysis procedure of ERP datasets [25,26]. The electrode sites (always a cluster of neighboring sites representative of the spatial distribution of the components) and time windows (representative of the temporal distribution of the components) were chosen based on literature and observation of grand-average (i.e., averaged across all participants) ERP waveforms obtained in the current study. When a sensory-level component had a discernible peak latency (the peak was clearly localized to a precise point in time) in each participant, peak latency was analyzed (see [25] for more on latency analysis).

The oddball effect was revealed as differences in the mean amplitude or peak latency of the standard and either the same- or different-category deviant. The category effect was seen by a differential response to the same- and different-category deviants. ANOVAs were used in the analyses, and Greenhouse–Geisser corrections were applied when sphericity assumption was violated. LSD tests were used for pairwise comparisons to follow up significant main effects. Preliminary analyses with color set as a factor revealed no significant Set  $\times$  Color interaction, all  $ps > 0.1$ . Therefore all ERP analyses are presented for data collapsed across sets. Figure 3 shows the ERP waveforms at eight representative electrode sites.

###### 1. Sensory-Level Component: P1

Figure 4 summarizes the results from the sensory-level components (P1 and the anterior N1). The P1 component was at its maximal amplitude at occipital electrode sites at around 150 ms after stimulus onset. But because the peak was not clearly discernible across individuals despite a recognizable P1 component in the grand-average waveforms, no peak latency analysis was conducted. Its mean amplitudes were measured between 130 and 170 ms averaged over five occipital electrode sites (O1, Oz, O2, O11h, O12h), then analyzed with

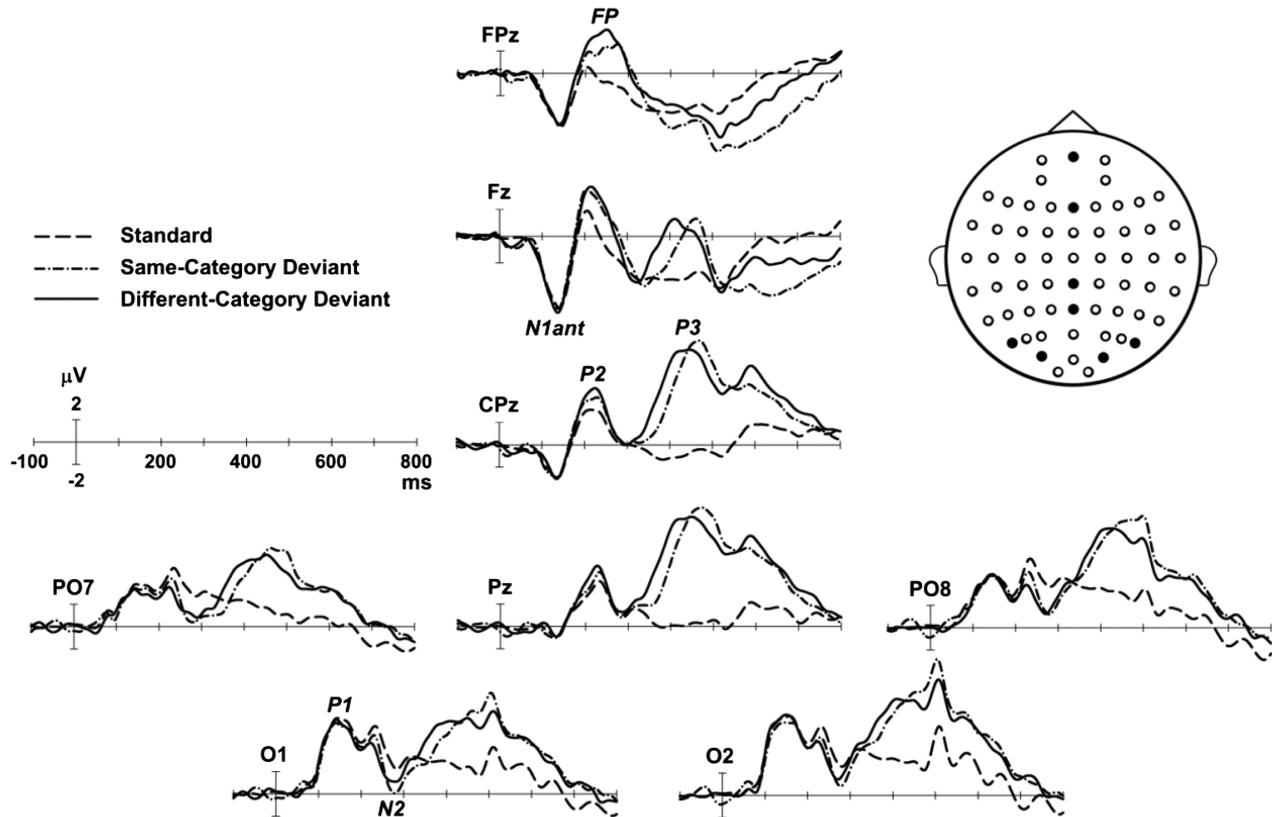


Fig. 3. Representative ERP waveforms at eight electrode locations, averaged for all participants ( $N = 15$ ). Each plot gives a different electrode location, and the electrode location is denoted above the y axis for each waveform plot (e.g., O1). The electrode location can be identified in the electrode location map (top right), where the demonstrated channels are marked as filled dots and where the spatial layout aligns with the spatial layout of the plots. The ERP components (e.g., P1) are labeled on one waveform each: N1ant denotes the anterior N1 component, and FP represents the frontal positivity.

a one-way ANOVA, with the only factor being Color (standard, same-category deviant, different-category deviant). The main effect did not approach significance,  $F(2, 28) = 1.19$ ,  $p = 0.32$ .

## 2. Sensory-Level Component: N1

The visual N1 component is usually divided into and separately analyzed as two subcomponents, namely the anterior N1 (around 140 ms at frontal to central electrode sites) and the posterior N1 (around 170 ms at occipital electrode sites) (e.g., [56]). We did not find the posterior N1 component in the current study. A strong anterior N1 component was maximal at midline fronto-central electrode sites around 140 ms post-stimulus and was analyzed as mean amplitudes between 120 and 160 ms over eight fronto-central locations (F1, Fz, F2, AF3, AF4, FC1, FCz, FC2). Again, the one-way ANOVA revealed no significant results,  $F(2, 28) = .06$ ,  $p = 0.94$ . Latencies were analyzed with a one-way ANOVA. The result did not reach significance either,  $F(2, 28) = 2.97$ ,  $p = 0.076$  (Fig. 4).

## 3. P2

There was a small P2 effect, which was found at centroparietal electrode sites after 200 ms post stimulus onset. This was quantified as mean amplitudes (204–224 ms) collapsed over four centroparietal electrode sites (CP1, CPz, CP2, Pz) and was analyzed with a one-way ANOVA. The main effect was significant,  $F(2, 28) = 4.81$ ,  $p = 0.016$ . However, the only difference we found with LSD pairwise comparisons was that

the different-category deviant color elicited larger amplitude than the same-category deviant color,  $p = 0.012$ . The amplitude of the same-category deviant condition did not differ from the other two conditions,  $ps > 0.10$  (Fig. 5).

## 4. N2

The N2 peak had a distribution over the lateral parieto-occipital electrode sites, starting at around 200 ms and lasting roughly 100 ms. Visual inspection of the ERP waveforms suggested that the deviants generally had stronger negative activities in contrast to the standard. Visual inspection of the ERP waveforms also suggested that the two deviants differed from each other in amplitude but within a narrow time window, indicating an evident but relatively time-limited category effect. Hence this component was analyzed as mean

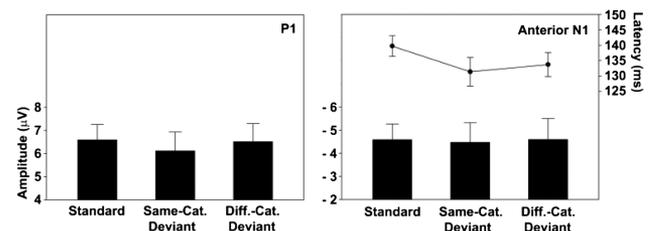


Fig. 4. Results from the sensory-level ERP components: mean amplitudes and peak latencies of P1 and the anterior N1 for standard and deviant colors. There are no significant differences between conditions in either amplitudes or latencies. Error bars indicate  $\pm$ SEs.

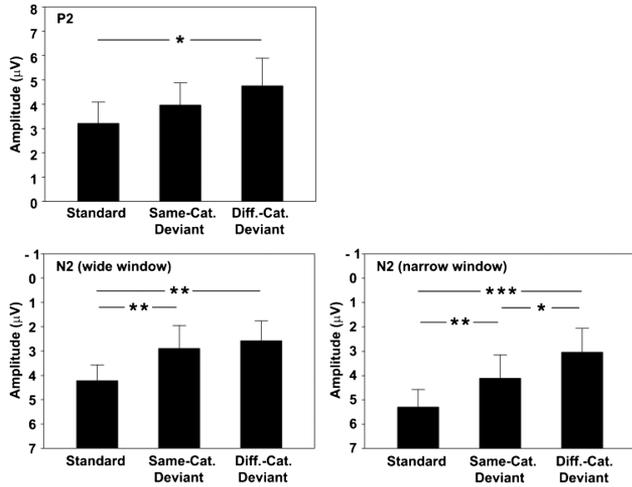


Fig. 5. Mean amplitudes of P2 and N2 components for standard and deviant colors. Asterisks indicate significant differences between conditions (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). Error bars indicate  $\pm$ SEs.

amplitudes over two time windows: a wide one (200–280 ms) and a narrow one (230–246 ms). This latter analysis was specifically to test whether the category effect takes place at the time range of the N2 component. Sites chosen for these analyses were O1, O2, PO7, PO5, PO6, PO8, OI1h, OI2h in the parieto-occipital areas. Unlike previously analyzed components, the N2 had a lateral distribution. Therefore a factor of hemisphere was added for the analysis, in which ERP activities were averaged across four electrode sites for the left and right hemisphere separately. The results are shown in Fig. 5 (collapsed across hemisphere).

For the wide-window data, we ran a two-way ANOVA: Color (standard, same-category deviant, different-category deviant)  $\times$  Hemisphere (left versus right). A significant main effect of Color was found,  $F(2, 28) = 10.33$ ,  $p = 0.00044$ . Subsequent LSD tests confirmed our observation by showing more negative amplitudes from the deviants than from the standard color,  $ps < 0.006$ . The amplitudes of the deviant conditions were not significantly different,  $p = 0.37$ . The effect of Hemisphere did not reach significance,  $F(1, 14) = 0.64$ ,  $p = 0.44$ , or interact with Color,  $F(2, 28) = 0.08$ ,  $p = 0.92$ .

Analysis for the narrow-window (230–246 ms) data demonstrated a slightly different pattern. The main effect of Color was significant,  $F(2, 28) = 13.76$ ,  $p < 0.0001$ . The activity was more negative for the different-category deviant condition than for the same-category deviant color,  $p = 0.020$ , and they were both stronger than the standard condition,  $p = 0.0083$  (same-category deviant) and  $p = 0.00042$  (different-category deviant). Similar to the wide-window data, Hemisphere did not have an effect on the amplitudes,  $F(1, 14) = 0.70$ ,  $p = 0.42$ , or interact with Color,  $F(2, 28) = 0.28$ ,  $p = 0.76$ .

### 5. Frontal Positivity

The three stimulus types elicited different activities over the frontal electrode sites. This effect started at around 200 ms post-stimulus and spanned more than 100 ms. Generally deviant colors produced a more positive deflection relative to the standard color. This effect was quantified as mean amplitudes over an early window (210–260 ms) and a late window

(280–320 ms) at frontal electrode sites FP1, FPz, FP2, AF3, and AF4.

Figure 6 summarizes the results of both analyses. The one-way ANOVA for the early-window data yielded a significant main effect of Color,  $F(2, 18) = 21.06$ ,  $p < 0.0001$ . LSD tests confirmed our visual inspection, showing that the deviants generated more positive activities than the standard color,  $ps < 0.0013$ , and that the activity was more positive for the different-category deviant color than the same-category deviant color,  $p = 0.042$ .

The analysis for the late-window data replicated the pattern that the deviant colors had more prominent positivity than the standard color,  $F(2, 28) = 10.54$ ,  $p = 0.00039$ , pairwise test  $ps < 0.0038$ . In contrast to the early-window results, there was no difference between the amplitudes of the two deviant conditions,  $p = 0.43$ .

### 6. P3

We found a long-lasting and strong P3 component at posterior (centro-parietal, parietal, parieto-occipital) electrode sites. As is common in visual oddball studies [17], it is evident that the deviant conditions had much larger P3 than the standard condition. This was confirmed by a two-way (Color  $\times$  Region) ANOVA, which statistically assessed the mean amplitudes over three regions, each of which included three electrode sites (centro-parietal: CP1, CPz, CP2; parietal: P1, Pz, P2; parieto-occipital: PO3, POz, PO4) between 350 and 530 ms. The main effect of Color was highly significant,  $F(2, 28) = 61.80$ ,  $p < 0.0001$ . Further paired LSD tests indicated that the deviants produced stronger activities than the standard,  $ps < 0.0001$ , suggesting that participants were actively engaged in the detection of the deviant stimuli and that the amplitudes of the deviants were similar,  $p = 0.56$  (Fig. 7). The effect of Region was also significant,  $F(2, 28) = 11.88$ ,  $p = 0.0031$ . This was because the activities were more prominent over the more posterior (parietal and parieto-occipital) sites than the centro-parietal sites,  $ps < 0.0032$  (parietal versus parieto-occipital  $p = 0.050$ ). The interaction did not reach significance,  $F(4, 56) = 3.34$ ,  $p = 0.052$ .

However, it cannot be concluded that the two deviant conditions elicited similar P3s, because we noticed in the waveforms that the two deviants had different patterns at different stages. Initially the amplitude of the different-category deviant condition was larger than that of the same-category deviant condition. Yet at a later stage this was reversed. Hence, we carried out two further ANOVAs to look into the P3 effects at these two stages (early 350–410 ms, late 460–520 ms). Because our aim was to test the differences between the two deviant conditions, we did not include the standard

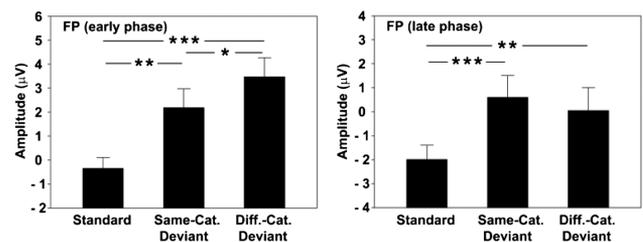


Fig. 6. Mean amplitudes of the frontal positivity (FP) for standard and deviant colors. Significant differences are marked by asterisks (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). Error bars indicate  $\pm$ SEs.

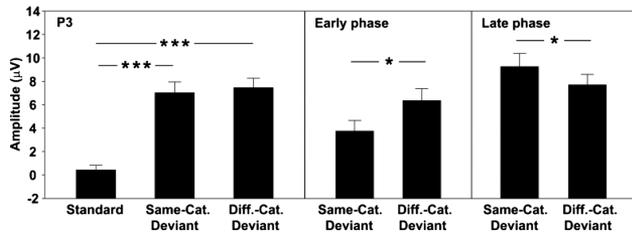


Fig. 7. Mean amplitude of the P3 component for standard and deviant colors. As indicated by asterisks ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ), deviant colors have a much larger P3, and the results on the same- and different-category deviants differ at different time windows. Error bars indicate  $\pm$ SEs.

condition in the analyses, resulting in 2 (Color)  $\times$  3 (Region) ANOVAs.

At the early stage, the main effect of Color was significant,  $F(1, 14) = 5.74$ ,  $p = 0.031$ , with the different-category deviant color eliciting a stronger P3. The amplitudes differed across regions,  $F(2, 28) = 10.01$ ,  $p = 0.0058$ , with larger P3 amplitude over the more posterior sites, all pairwise  $ps < 0.039$ . The interaction was not significant,  $F(2, 28) = 1.79$ ,  $p = 0.20$ . For the late stage, an opposite pattern was found, with the same-category deviant eliciting stronger P3 activities than the different-category deviant condition,  $F(1, 14) = 5.80$ ,  $p = 0.030$ . Effect of Region was again significant,  $F(2, 28) = 11.86$ ,  $p = 0.0028$ , because amplitudes were larger at the parietal and parieto-occipital sites than at the centroparietal sites,  $ps < 0.0037$ , but did not differ between the parietal and parieto-occipital regions,  $p = 0.23$ . No significant interaction was found,  $F(2, 28) = 0.51$ ,  $p = 0.50$ .

## 5. DISCUSSION

The aim of the current study was to investigate the time course of category effects for equally discriminable same- and different-category colors. Same- and different-category colors were equated in discriminability in terms of the number of JNDs. The category effect was then assessed using a visual oddball task where infrequent deviant colors were either from the same- or different-color category to a frequent standard color. Although there was no category effect on a behavioral version of this task, the ERPs revealed clear category effects for several ERP components. ERP components elicited in response to infrequent deviant colors varied according to whether deviant colors were from the same- or different-color category to the frequent standard color. This category effect was only found after the sensory-level processing stage, after 200 ms post stimulus presentation.

Different-category deviants elicited greater mean amplitude than same-category deviants, within a limited time window of the N2 component (230–246 ms) at occipital sites, and during the early phase of a frontal positivity (210–260 ms). For the P3 component at parietal sites, we also found an amplitude enhancement for the different-category deviant relative to the same-category deviant at the early phase (350–410 ms) and a reduction at the late phase (460–520 ms). This pattern actually suggests that the P3 component reached its amplitude maximum earlier for the different-category deviants than for the same-category deviants (see Figs. 3 and 7), suggesting an earlier context updating process for the different-category deviant stimuli [43]. Importantly however, a category effect was absent from ERP activities earlier than 200 ms. Two early ERP

components (P1 and the anterior N1) were observed in the current study, and there was no trace of a category effect for either of these components.

Although many prior ERP investigations of color categories found effects early on in the time course [16–22], the category effects in those studies cannot be dissociated from inequalities in the color space used to equate same- and different-category colors. In contrast to previous studies, the category effects found here cannot be due to differences in the discriminability of same- and different-category colors as colors were equated in empirically measured JNDs. Other stimulus issues (such as salience of one hue over the others) cannot account for the category effects, as the design ensured that different- or same-category deviant hues varied across participants (due to the use of two color sets and individual differences in color naming). Therefore, we are confident that the differences in ERP components for different- and same-category deviants is not due to stimulus issues, but is due to the categorical relationship of the deviant and the standard.

Here, we show that when same- and different-category colors are equated in their discriminability, the category effects at early perceptual stages of processing disappear and only post-perceptual category effects remain. These findings align with a prior ERP study of newly trained color categories, which was able to dissociate category effects from stimulus issues and which only found effects later on in the time course [15]. Category effects for the blue–green categorical distinction in the current study were found earlier than for the newly trained categories in the prior study (e.g., from 210 ms for blue–green versus 350 ms for the newly trained category boundary in Clifford *et al.*) [15]. This potentially suggests that the degree to which the categorical distinction is established could influence the time course of the category effect.

The findings of the current study challenge the idea that cross-linguistic differences in color terms would result in pre-attentive differences in perceptual processing [19]. If the basic color categories of blue and green do not affect perceptual processing, then it is also unlikely that cross-linguistic differences in color lexicons would result in differences in early perceptual processing of color. It has previously been argued that the domain of color provides a prime example of Whorfian effects of language on perceptual representation [57]. However, the findings of the current study suggest that color terms in fact have a less pervasive influence on perception than such accounts propose, and although color categories interact with post-perceptual processes, the perceptual representation of color is actually unaffected.

## 6. CONCLUSION

The current study finds that color categories only affect post-perceptual color processing when same- and different-category colors are equally discriminable. It therefore appears that color categories do affect people's ability to distinguish colors [53], but that this is due to an influence of categories on post-perceptual processes rather than due to categories affecting the perceptual similarity of colors or interacting with their perceptual representation.

## ACKNOWLEDGMENTS

The research was supported by a European Research Council Starter Grant (project "CATEGORIES," ref. 283605) to AF, and

CW was supported by a DAAD postdoctoral fellowship. We would like to thank Amanda Holmes for contributing to the development of the visual oddball task and for insightful discussion, and two anonymous reviewers for highly constructive comments.

## REFERENCES

1. M. R. Pointer and G. G. Attridge, "The number of discernible colours," *Color Res. Appl.* **23**, 52–54 (1998).
2. J. M. M. Linhares, P. D. Pinto, and S. M. C. Nascimento, "The number of discernible colors in natural scenes," *J. Opt. Soc. Am. A* **25**, 2918–2924 (2008).
3. P. Kay and T. Regier, "Language, thought and color: recent developments," *Trends Cogn. Sci.* **10**, 51–54 (2006).
4. C. E. Sternheim and R. M. Boynton, "Uniqueness of perceived hues investigated with a continuous judgmental technique," *J. Exp. Psychol.* **72**, 770–776 (1966).
5. S. R. Harnad, "Interhemispheric division of labour," presented at *Bucke Society Conference on Transformation of Consciousness*, Montreal, 1973.
6. A. L. Gilbert, T. Regier, P. Kay, and R. B. Ivry, "Whorf hypothesis is supported in the right visual field but not in the left," *Proc. Natl. Acad. Sci. USA* **103**, 489–494 (2006).
7. G. V. Drivonikou, P. Kay, T. Regier, R. B. Ivry, A. L. Gilbert, A. Franklin, and I. R. Davies, "Further evidence that Whorfian effects are stronger in the right visual field than the left," *Proc. Natl. Acad. Sci. USA* **104**, 1097–1102 (2007).
8. C. A. Daoutis, M. Pilling, and I. R. Davies, "Categorical effects in visual search for colour," *Vis. Cogn.* **14**, 217–240 (2006).
9. D. Roberson and J. Davidoff, "The categorical perception of colors and facial expressions: the effect of verbal interference," *Mem. Cogn.* **28**, 977–986 (2000).
10. P. Kay and W. Kempton, "What is the Sapir-Whorf hypothesis," *Am. Anthropol.* **86**, 65–79 (1984).
11. D. Roberson and J. R. Hanley, "Color vision: color categories vary with language after all," *Curr. Biol.* **17**, R605–R607 (2007).
12. M. A. Webster and P. Kay, "Color categories and color appearance," *Cognition* **122**, 375–392 (2012).
13. K. T. Mullen and J. J. Kulikowski, "Wavelength discrimination at detection threshold," *J. Opt. Soc. Am. A* **7**, 733–742 (1990).
14. J. Krauskopf, D. R. Williams, M. B. Mandler, and A. M. Brown, "Higher order color mechanisms," *Vis. Res.* **26**, 23–32 (1986).
15. A. Clifford, A. Franklin, A. Holmes, V. G. Drivonikou, E. Özgen, and I. R. Davies, "Neural correlates of acquired color category effects," *Brain Cogn.* **80**, 126–143 (2012).
16. E. Fonteneau and J. Davidoff, "Neural correlates of colour categories," *Neuroreport* **18**, 1323–1327 (2007).
17. A. Holmes, A. Franklin, A. Clifford, and I. R. Davies, "Neurophysiological evidence for categorical perception of color," *Brain Cogn.* **69**, 426–434 (2009).
18. Q. Liu, H. Li, J. L. Campos, Q. Wang, Y. Zhang, J. Qiu, Q. Zhang, and H. J. Sun, "The N2pc component in ERP and the lateralization effect of language on color perception," *Neurosci. Lett.* **454**, 58–61 (2009).
19. G. Thierry, P. Athanasopoulos, A. Wiggett, B. Dering, and J. R. Kuipers, "Unconscious effects of language-specific terminology on preattentive color perception," *Proc. Natl. Acad. Sci. USA* **106**, 4567–4570 (2009).
20. A. Clifford, A. Holmes, I. R. Davies, and A. Franklin, "Color categories affect pre-attentive color perception," *Biol. Psychol.* **85**, 275–282 (2010).
21. L. Mo, G. Xu, P. Kay, and L. H. Tan, "Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color," *Proc. Natl. Acad. Sci. USA* **108**, 14026–14030 (2011).
22. A. Clifford, A. Franklin, I. R. Davies, and A. Holmes, "Electrophysiological markers of categorical perception of color in 7-month old infants," *Brain Cogn.* **71**, 165–172 (2009).
23. P. Athanasopoulos, "Cognitive representation of colour in bilinguals: the case of Greek blues," *Bilingual. Lang. Cogn.* **12**, 83–95 (2009).
24. Q. Liu, H. Li, J. L. Campos, C. Teeter, W. Tao, Q. Zhang, and H. J. Sun, "Language suppression effects on the categorical perception of colour as evidenced through ERPs," *Biol. Psychol.* **85**, 45–52 (2010).
25. S. J. Luck, *An Introduction to the Event-Related Potential Technique* (MIT, 2005).
26. M. D. Rugg and M. G. H. Coles, *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition* (Oxford University, 1996).
27. I. Czigler, "Visual mismatch negativity: violation of nonattended environmental regularities," *J. Psychophysiol.* **21**, 224–230 (2007).
28. I. Czigler, L. Balázs, and L. G. Pató, "Visual change detection: event-related potentials are dependent on stimulus location in humans," *Neurosci. Lett.* **364**, 149–153 (2004).
29. W. T. Siok, P. Kay, W. S. Wang, A. H. Chan, L. Chen, K. K. Luke, and L. H. Tan, "Language regions of brain are operative in color perception," *Proc. Natl. Acad. Sci. USA* **106**, 8140–8145 (2009).
30. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, 2000).
31. R. W. G. Hunt and M. R. Pointer, *Measuring Colour*, 4th ed. (Wiley, 2011).
32. A. M. Brown, D. T. Lindsey, and K. M. Guckes, "Color names, color categories, and color-cued visual search: sometimes, color perception is not categorical," *J. Vis.* **11**(12), 2 (2011).
33. D. Roberson, I. R. Davies, and J. Davidoff, "Color categories are not universal: replications and new evidence from a stone-age culture," *J. Exp. Psychol. Gen.* **129**, 369–398 (2000).
34. C. A. Daoutis, A. Franklin, A. Riddett, A. Clifford, and I. R. Davies, "Categorical effects in children's colour search: a cross-linguistic comparison," *Br. J. Dev. Psychol.* **24**, 373–400 (2006).
35. J. Winawer, N. Witthoft, M. C. Frank, L. Wu, A. R. Wade, and L. Boroditsky, "Russian blues reveal effects of language on color discrimination," *Proc. Natl. Acad. Sci. USA* **104**, 7780–7785 (2007).
36. E. Özgen and I. R. Davies, "Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis," *J. Exp. Psychol. Gen.* **131**, 477–493 (2002).
37. C. Witzel, T. Hansen, and K. R. Gegenfurtner, "Categorical reaction times for equally discriminable colours," *Perception* **38**, 14 (2009).
38. C. Witzel and K. R. Gegenfurtner, "Category effects for red and brown," presented at *European Conference of Visual Perception (ECVP)*, Alghero, Italy, 2012.
39. C. Witzel and K. R. Gegenfurtner, are preparing a manuscript to be called "Categorical facilitation effects for equally discriminable colours."
40. S. Johannes, T. F. Münte, H. J. Heinze, and G. R. Mangun, "Luminance and spatial attention effects on early visual processing," *Cogn. Brain Res.* **2**, 189–205 (1995).
41. S. A. Hillyard, E. K. Vogel, and S. J. Luck, "Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence," *Phil. Trans. R. Soc. London B* **353**, 1257–1270 (1998).
42. A. Martínez, F. Di Russo, L. Anillo-Vento, M. I. Sereno, R. B. Buxton, and S. A. Hillyard, "Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas," *Vis. Res.* **41**, 1437–1457 (2001).
43. E. Donchin, "Surprise! ... surprise?" *Psychophysiology* **18**, 493–513 (1981).
44. G. McCarthy and E. Donchin, "A metric for thought: a comparison of P300 latency and reaction time," *Science* **211**, 77–80 (1981).
45. S. H. Patel and P. N. Azzam, "Characterization of N200 and P300: selected studies of the event-related potential," *Int. J. Med. Sci.* **2**, 147–154 (2005).
46. S. J. Luck and S. A. Hillyard, "Spatial filtering during visual search: evidence from human electrophysiology," *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 1000–1014 (1994).
47. S. J. Luck and S. A. Hillyard, "Electrophysiological correlates of feature analysis during visual search," *Psychophysiology* **31**, 291–308 (1994).

48. S. Ishihara, *Ishihara Tests for Color-Blindness* (Kanehara, 1987).
49. R. Fletcher, *City Color Vision Test* (Keeler, 1981).
50. D. H. Brainard, "The psychophysics toolbox," *Spat. Vis.* **10**, 433–436 (1997).
51. D. G. Pelli, "The VideoToolbox software for visual psychophysics: transforming numbers into movies," *Spat. Vis.* **10**, 437–442 (1997).
52. C. Witzel and A. Franklin, "Do focal colors look particularly 'colorful'?" *J. Opt. Soc. Am. A* (to be published).
53. C. Witzel and K. R. Gegenfurtner, "Categorical sensitivity to color differences," *J. Vis.* **13**(7), 1 (2013).
54. R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.* **112**, 713–719 (2001).
55. E. Niedermeyer, "Alpha rhythms as physiological and abnormal phenomena," *Int. J. Psychophysiol.* **26**, 31–49 (1997).
56. H. J. Heinze, S. J. Luck, G. R. Mangun, and S. A. Hillyard, "Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection," *Electroen. Clin. Neuro.* **75**, 511–527 (1990).
57. G. Lupyan, "Linguistically modulated perception and cognition: the label-feedback hypothesis," *Front. Psychol.* **3**, 1–13 (2012).