Cone Excitation Ratios Correlate With Color Discrimination Performance in the Horse (Equus caballus)

Carol A. Hall
Nottingham Trent University

Helen J. Cassaday and Chris J. Vincent
University of Nottingham

Andrew M. Derrington
Newcastle University

Six horses (Equus caballus) were trained to discriminate color from grays in a counterbalanced sequence in which lightness cues were irrelevant. Subsequently, the pretrained colors were presented in a different sequence. Two sets of novel colors paired with novel grays were also tested. Performance was just as good in these transfer tests. Once the horse had learned to select the chromatic from the achromatic stimulus, regardless of the specific color, they were immediately able to apply this rule to novel stimuli. In terms of the underlying visual mechanisms, the present study showed for the first time that the spectral sensitivity of horse cone photopigments, measured as cone excitation ratios, was correlated with color discrimination performance, measured as accuracy, repeated errors, and latency of approach.

**Keywords:** horse, color vision, cone excitation ratios, discrimination learning

An animal is said to possess color vision if it can discriminate between stimuli that differ in their distributions of spectral energy, independent of any differences in total energy (Jacobs, 1993). Although many animals in most phyla have been shown to have some color vision (Kelber, Vorobyev, & Osorio, 2003), the extent of these abilities and how they differ from one animal to the next requires further investigation. In particular, normal human color vision provides no basis to predict what stimuli will be salient to the horse. For example, one consequence of dichromacy is that a wavelength will exist within the spectrum that is confusable with white, called the dichromatic neutral point. This is generally located within the middle wavelengths, for example, at around 505 nm (green) in the tree shrew (Jacobs & Neitz, 1986). This point has been located at about 480 nm in the horse (Geisbauer, Griebel, Schmid, & Timney, 2004).

Research into horses’ ability to see colors has resulted in conflicting findings, both in studies adopting a behavioral approach (Grzimek, 1952; Macuda & Timney, 1999; Pick, Lovell, Brown, & Dail, 1994; Smith & Goldman, 1999) and in those investigating the spectral sensitivity of the cone photopigments of the equine retina (Carroll, Murphy, Neitz, Ver Hoeve, & Neitz, 2001; Macuda, 2000; Ver Hoeve, Neitz, & Murphy, 1999; Yokoyama & Radlwimmer, 1999). There is evidence that the horse, like other ungulates, possesses two types of cone photopigment, one maximally sensitive to short wavelengths, the other maximally sensitive to medium-long wavelengths. Values for the spectral peaks of these photopigments have been estimated at 545 nm (Macuda, 2000; Ver Hoeve et al., 1999; Yokoyama & Radlwimmer, 1999) and 429 nm (Macuda, 2000). These values vary slightly from those (539 nm and 428 nm) obtained by Carroll et al. (2001), but both are consistent with the range found in other ungulates (Jacobs, 1993). The former values were used in the current study.

Behavioral studies into the ability of animals to see colors have generally involved training them to discriminate between chromatic and achromatic stimuli, where all other cues (in particular differences in lightness/brightness and olfactory or spatial cues) have been made irrelevant. The length of time required for this training has limited both the number of subjects used and the range of colors that have been tested by this method. So far, a maximum of four colors have been tested on the horse. Although all previous studies have concluded that this species has the ability to discriminate blue from gray, the results for red, green, and yellow have varied. Grzimek (1952) found that horses could discriminate yellow best, followed by green and then blue, but had some difficulty with red. Horses tested by Pick et al. (1994) could reliably discriminate blue and red, but not green, from gray. Similar results were obtained by Macuda and Timney (1999), whose horses could discriminate red and blue, but not green or yellow, from gray (although, with the green stimulus used in their study, they noted that there was evidence that at least some of the chromatic information was available).

A study into color vision in another ungulate, the fallow deer (Dama dama), demonstrated that this species could discriminate greens from grays (Birgersson, Alm, & Forkman, 2001). Given the conformity that has been found in the peak sensitivities of the
One notable difference between the early study by Grzimek (1952) and the more recent investigations (Macuda & Timney, 1999; Pick et al., 1994; Smith & Goldman, 1999) is the height at which the stimuli were presented. Successful discrimination of greens was demonstrated when the stimuli were presented at ground level both in horses (Grzimek, 1952) and deer (Birgersson et al., 2001) but not when they were presented to horses at nose height (Macuda & Timney, 1999; Pick et al., 1994). Smith and Goldman (1999) presented their stimuli at a height of 1.22 m from the ground and found individual differences in the performance of the horses. We have found that presenting stimuli at ground level enhanced the performance of horses in a task of lightness discrimination (Hall, Cassaday, & Derrington, 2003). Thus, stimulus position is likely to have been a contributing factor to the discrepant findings of the previous studies.

The visual acuity of the horse has been found to vary according to the color of the stimulus, being poorer with blue targets than with other colors, including yellow (Grzimek, 1952). The visual acuity value obtained by Grzimek (1952) for the blue stimulus corresponded closely to that obtained for the peripheral retina (2.7 cycles per degree) and that for the yellow stimulus was closer to the value obtained for the visual streak (16.5 cycles per degree; Harman, Moore, Hoskins, & Keller, 1999). Therefore, to guarantee the visibility of all of the colors presented, they should be of a size that, when presented from the assigned distance, results in a visual angle that is greater than 0.5°.

As information relating to the spectral sensitivity of the cone photopigments of the horse is now available, data obtained from behavioral methods can also be compared with the predicted effect of colors on the equine visual system. Using the values for the spectral peaks of the cone photopigments of the horse (429 and 545 nm; Macuda, 2000), spectral sensitivity curves can be derived (Govardovskii, Fyhrquist, Reuter, Kuzmin, & Donner, 2000). These can then be used to predict how easily the horse should be able to discriminate individual colors from gray.

The aim of this study was to train horses to discriminate a range of different colors from grays. Lightness cues were made irrelevant by mounting the colors randomly within a geometric pattern of mixed grays, similar to those used by von Frisch (1950) when investigating the color sense of bees. The same geometric pattern of mixed grays, but without the color, was used as the negative stimulus. Stimuli were presented at ground level to optimize discrimination accuracy (Hall et al., 2003). The training phase suggested that horses’ discrimination learning performance was based on chromatic information. Transfer tests, which presented familiar colors in a different sequence as well as novel colors, were used to confirm that the discrimination was learned on the basis of chromatic information. Finally, the reflectance spectrum of each of the colors was measured and cone excitation ratios were calculated. The cone excitation ratios for the colors and the grays were compared, and the difference between them was used as an indication of the effect that each color had on the visual system of the horse. Thus, the effect of the colored stimuli on the photopigments of the horse could be compared with the behavioral data. A reliable correlation between the cone excitation ratios and the discrimination performance would, in principle, provide a means of predicting the appearance of colors to the horse.

### Method

#### Subjects

The six subjects used were riding horses (*Equus caballus*) from the Brackenhurst Equestrian Centre, Nottingham Trent University, where the study was carried out. They consisted of four geldings and two mares whose ages ranged from 5 to 18 years, with a mean of 11.5 years. Their heights ranged from 1.52 to 1.65 m, with a mean of 1.58 m. During the period that the study was carried out, all of the horses were ridden approximately 6 days a week for 2 hr a day. All of the horses had previously been trained to perform a two-choice black–white discrimination task to a criterion of 70% accuracy on four consecutive training sessions (Hall et al., 2003). The same test area and method of displaying the stimuli and delivering the reward were used in the current study.

A veterinary surgeon examined the eyes of all six horses prior to the current study and found no evidence of any ocular abnormality. When taking part in the study, the horses all wore head collars from which the lead ropes had been removed.

#### Test Area and Apparatus

The test area was located in one half of an enclosed barn with a concrete floor, which was fenced off along the long side using galvanized wire mesh barriers (1.2 m in height) and screened from view by sheeting to a height of 3 m. A gap of 100 mm in this screening allowed the experimenter to view the subject performing the trials while remaining outside the test area. All of the training and testing was carried out in daylight, which illuminated the test area via skylights in the roof. The test area was 5 m wide and 10 m long; a “starting” line of masking tape was placed on the floor 6.5 m from the end wall where the stimuli were displayed.

Two identical wooden boxes were placed on the floor against the wall, each 1.25 m from the sidewall with a gap between the two boxes of 1.5 m. The top flap of the box sloped forward at an angle of 60° from the vertical and was hinged at the top to open inward. Perspex sheets were mounted on the opening flaps of each box, behind which the stimulus cards could be slotted. The box was designed to allow the top to be left locked or unlocked, without any visual cue to its state. The horses could open the box by pushing the flap with their noses. The flap of the box displaying the positive stimulus was left unlocked, whereas the flap of the box displaying the negative stimulus was locked. A correct choice was rewarded by access to the food within the box. This consisted of a small piece of carrot, approximately 30 × 10 × 10 mm, placed in each of the stimulus boxes so that olfactory cues could not guide stimulus selection. During training and testing, both boxes were treated identically with respect to changing the stimulus cards, opening and shutting the flaps, and removing or inserting the locking block, so auditory cues could not guide stimulus selection. For further details of the test area and design and dimensions of the stimulus boxes, see Hall et al. (2003).

#### Stimulus Cards

A range of 15 colors was produced for training and a further 6 colors were produced for the transfer tests. In both cases, horses were required to select chromatic from achromatic stimuli, where lightness could not be used as a cue. The stimulus cards measured 375 mm square and were subdivided into a regular geometric pattern of 8 different grays within which was mounted a colored panel in the case of the chromatic stimuli, or not in the case of the achromatic stimuli. The gray and colored stimuli were printed, cut to size, and glued on a piece of card. The whole card was then laminated to prevent damage when being inserted behind the Perspex panel.
on the front of the stimulus box. See Figure 1 for the design and dimensions of the stimulus cards.

The visual angle of the card was calculated using the trigonometric functions of the Casio scientific calculator (fx-83WA). It was found to be approximately 3° when presented from a distance of 6.5 m, that of the individual panels being approximately 1°. Both measurements were greater than the estimated minimum angle of 0.5° required for their visibility by the horse.

Four different arrangements of the eight different grays could be displayed in either vertical orientation, resulting in eight different configurations on the achromatic cards. For each of the colors tested, a set of four different stimulus cards was constructed, which could also be displayed in either vertical orientation, thus allowing the colored panel to appear in any of the eight positions within the card. The other seven positions contained seven of the same grays as those on the distracter cards.

Colors

The colored stimuli were produced using the computer graphics program Paint Shop Pro 5 (version 5.01). A range of 15 colors was produced for the initial training and a further 6 colors for the final test. The stimulus colors were printed using a Hewlett Packard Laser Jet 4500 color printer with a resolution of 600 × 600 dots per inch (dpi).

The 15 training colors were selected from across the spectrum and were presented in a sequence that related to their appearance to the trichromatic human. They were arranged according to their spectral appearance and presented to the subjects in one of two counterbalanced orders (either starting with the short or long wavelength colors). The colors were numbered C1 through C15. C1 was at the short wavelength end of the spectrum and C15 at the long wavelength end. Discrimination ability with 6 novel colors was tested in the same way in the transfer tests.

Grays

The grays were produced using the grayscale option of the same computer graphics program and a black-and-white laser printer (Hewlett Packard, 4000 N). The vertical resolution rating of the printer was 600 dpi, the same as the color printer. Only 256 different shades of gray are available in Paint Shop Pro 5 (version 5.01), resulting in a reduction in variation once colors are converted to grayscale. This meant that several colors had the same value in grayscale. The perceived lightness of the colors would depend on the spectral sensitivity of the visual system and may not be equivalent to the match generated by the graphics program. However, given the random location of the colors within 7 different shades of gray, the possibility that these lightness cues could still be used effectively was minimal.

Spectral Reflectance of the Colors and Grays

The spectral reflectance of each of the colors and grays when mounted on the stimulus cards was measured under the experimental conditions using a Minolta CS1000 spectrophotometer. The reflectance measurements were recorded under the same illumination and with the stimuli in the same position as in the subsequent discrimination training. Spectral radiance values (radians) for wavelengths between 380 and 780 nm were recorded in 1-nm steps. To minimize the effect of minor variations in environmental conditions, we took two separate readings for each color at each wavelength and calculated the mean value. The latter was then used in subsequent calculations. The data were downloaded onto Excel files, and line graphs to show the reflectance spectrum for each of the colors were then plotted. Given the range of light visible to the human and the spectral sensitivity of the horse’s photoreceptors, only the radiance values for wavelengths of between 380 and 720 nm were used. The reflectance spectrum of each of the training colors and of the darkest and lightest gray is shown in Figures 2, 3, and 4.

Design

Color discrimination training. The horses were divided into two groups, each consisting of two geldings and one mare, that were tested in one of two counterbalanced sequences. The first group started their training with C1 and the second group with C15. The performance of each subject for each of the 15 colors was assessed using four separate measures: number of trials to criterion, accuracy (the percentage of correct choices made at the first attempt), errors (the percentage of repeated incorrect choices made at each of the trials during the session), and latency of approach to the stimulus (measured in seconds). The approach to the stimulus was timed from the release of the horse at the starting line to when the stimulus box was touched. Latency of approach to each color was calculated separately for correct and incorrect choices. A comparison of performance was made using the correct choice latencies for each color. These scores were then compared with the scores for incorrect choices.

Transfer tests. The 15 pretrained colors were then randomly arranged into 5 sets (S1–S5), each containing 3 of the colors. The learning criterion for each set was the same as during the training sessions (10 consecutive correct choices within one session); once this had been reached for all 5 sets, the novel color sets were tested.

The six novel colors (that were mounted within the novel grays and paired with the novel achromatic stimuli) were arranged into two sets (NS1 and NS2); NS1 consisted of a novel green (NC1), yellow (NC2), and brown (NC3); NS2 consisted of a novel blue (NC4), purple (NC5), and pink (NC6). The ability of the horses to discriminate these novel sets of colors was then tested. The first group started with NS1; the second group started with NS2.

Procedure

Color discrimination training. Color discrimination training sessions were carried out two to three times/week and consisted of 16 presentations of the paired stimuli (each presentation = 1 trial). The left–right position of the positive stimulus was varied randomly, up to a maximum of three consecutive choices on one side, to avoid spatial cues from becoming more important than visual cues. Within each training session, equal numbers of left and right presentations of the positive stimulus were included. The design of the stimulus cards meant that the horses could not use brightness
Figure 2. The reflectance spectrum of each of the training colors C1–C6 used in the color discrimination training, as obtained using a Minolta CS1000 spectroradiometer. The relative proportions of the different wavelengths of light (between 380 and 720 nm) that were reflected by the stimuli under the experimental conditions are shown. Values on the x-axis are the wavelengths of light (nm). Values on the y-axis refer to the radiance at each of the measured wavelengths (radians).
Figure 3. The reflectance spectrum of each of the training colors C7–C12 used in the color discrimination training, as obtained using a Minolta CS1000 spectroradiometer. The relative proportions of the different wavelengths of light (between 380 and 720 nm) that were reflected by the stimuli under the experimental conditions are shown. Values on the x-axis are the wavelengths of light (nm), Values on the y-axis refer to the radiance (radians).
Figure 4. The reflectance spectrum of each of the training colors C13–C15 and the darkest and lightest gray (G1 and G8) used in the color discrimination training, as obtained using a Minolta CS1000 spectroradiometer. The relative proportions of the different wavelengths of light (between 380 and 720 nm) that were reflected by the stimuli under the experimental conditions are shown. Values on the x-axis are the wavelengths of light (nm). Values on the y-axis refer to the radiance at each of the measured wavelengths (radians).
cues to make the discrimination. The stimuli were always presented at the same angle and were all illuminated from above.

At the start of each session, the horse was led into the barn, the doors were closed, and the horse was positioned behind the starting line, directly facing the stimulus boxes. It was released by the handler (changing side from trial to trial in a roughly random order to control for directional influences) and allowed to approach the boxes to make its selection. A correct choice was rewarded by access to the carrot via the unlocked flap before the subject was caught and led behind the screens. The number of trials on which the horse made a correct selection at the first attempt was calculated as a percentage of the total number of trials and resulted in an accuracy score. An incorrect choice resulted in the horse being caught by the handler before it could try the correct box and being led back to the starting line to try again. If an incorrect choice was made, the same presentation was repeated until the horse made the correct choice. The initial choice would be scored incorrect, and repeated errors with the same stimulus presentation were summed within any one trial. The number of error runs (on first or subsequent attempts) was calculated as a percentage of the total number of runs (whether correct or incorrect) in that session. Thus, the error scores were not simply the obverse of accuracy scores but rather reflected the horse’s tendency to persist in making an incorrect choice.

At the end of each trial, the horse was led behind the screens while the experimenter repositioned the stimuli according to the rearranged semi-random order. When no change of stimulus position was required, the cards were removed and replaced in the same box to control for possible auditory cues. After a period of 30 s, the horse was led back to the starting line to commence the next trial.

Training with each color continued until a learning criterion of 10 consecutive correct choices had been made within one training session; the next color was then introduced in the following session. The probability of making 10 correct choices by chance was $p < .001$, and training with each color continued until this criterion had been reached.

**Transfer tests.** The basic test procedure was exactly the same as that used in training. Controls for spatial cues and the random positioning of the colored panel within the positive stimulus were the same. Each session consisted of up to 20 trials, with the colors appearing randomly (each color would be presented at least five times during each session unless the learning criterion had been reached prior to that; colors would not be repeated more than twice consecutively). The same measures of performance that had been used in training were recorded throughout the testing phase (trials to criterion, accuracy, errors, and latency to approach). The colors included in each set and the order in which they were tested are shown in Table 1.

**Cone excitation values of the colored stimuli.** The spectral sensitivity of an A$_1$-based photopigment was modeled using the process detailed in Govardovski et al. (2000) for the peak sensitivities of the photoreceptors of the horse (429 and 545 nm; Macuda, 2000). The nomogram produced by this process gives the spectral sensitivity of the cones over the same range of wavelengths as used to measure the spectral radiance of the colored stimuli. The resultant cone fundamental spectral sensitivity curves were produced in Matlab, using the Psychtoolbox GavardovskiNomogram function (Brainard, 1997).

The function used a polynomial approximation to model photopigment spectral sensitivities consisting of alpha and beta band templates (Equations 1 and 2 in Govardovskii et al., 2000). The resultant approximation was combined with the spectral data for each stimulus to produce excitation values using standard techniques (Brainard, 1996).

To assess the effect that each of the colored stimuli had on each type of cone photopigment, we multiplied the spectral energy values of the former by the sensitivity values of the latter. In Matlab, the Psychtoolbox EnergyToCones function (Brainard, 1997) was applied to the energy value for each wavelength in discrete steps, resulting in cone excitation values at every wavelength. This series of values was integrated to obtain the cone excitation value for each stimulus.

The EnergyToCones function, therefore, took parameters relating to wavelength, spectral energy at a given wavelength, the sampling distribution used to represent data, and photopigment sensitivity. It returned cone excitations by looking up the appropriate sensitivity values and multiplying these with the given spectral value.

To obtain the relative effect of each color on the two cone photopigments of the horse, the cone excitation values were plotted on a graph in Excel (values on the x-axis for the M–L cones and on the y-axis for the S cones), and the ratio of y/x was calculated. The cone excitation values for the achromatic (grays, black, and white) stimuli were also plotted, and the ratio of y/x was calculated for each stimulus. The ratios were consistent to two decimal places, and the mean was calculated for use in subsequent calculations ($y = 0.3354x$). The difference from the achromatic excitation ratio was calculated for each of the chromatic stimuli, which was then used as a measure of the perceived chromaticity of the stimulus represented by that point. Cone excitation ratios for each of the training colors and the constant ratio for the achromatic stimuli are shown in Figure 5.

Because of possible differences in the contribution to the perception of brightness made by each of the two cone types (Stockman & Sharpe, 2000), those values that represented relatively greater excitation of the M–L cones were considered separately from those that represented greater excitation of the S cones. The correlation between these ratio differences and the behavioral scores was then investigated.

**Data Analysis**

**Color discrimination training.** The ability of the horses to discriminate each of the 15 colors was assessed by calculating the mean number of trials taken to reach the learning criterion (10 consecutive correct choices, $p < .001$), the mean accuracy and error percentages, and mean latency of approach. A Kolmogrov–Smirnov (Lilliefors) test was carried out in SPSS 9.0 for Windows to evaluate whether the data were normally distributed. The distribution of the accuracy and error percentages did not vary significantly from the normal distribution (accuracy $= 0.119$, $df = 15$, $p = .2$; errors $= 0.19$, $df = 15$, $p = .148$), but the distribution of the number of trials to criterion and latencies did (number of trials to criterion $= 0.269$, $df = 15$, $p = .005$; latency $= 0.226$, $df = 15$, $p = .038$). Consequently, although parametric tests were used to analyze the accuracy and error percentages, nonparametric equivalents were used for the trials to criterion and latencies.

The effect of color on the four measures of performance was assessed using the one-way repeated measures analysis of variance (ANOVA) for the parametric data and the Friedman test for the nonparametric data.

A comparison between the two groups of subjects was made to assess the possible effect that the order of presentation of the colors had on their training. The mean scores for each measure were calculated and compared using the t-test (unrelated) and the Mann–Whitney U test. All analyses were two-tailed.

---

**Table 1**

**Component Colors and Order of Presentation of the Color Sets Used in the Transfer Tests**

<table>
<thead>
<tr>
<th>Color set</th>
<th>Component colors</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>C4, C9, C14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td>C5, C10, C13</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>S3</td>
<td>C6, C11, C12</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>S4</td>
<td>C2, C8, C15</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>S5</td>
<td>C1, C5, C7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>NS1</td>
<td>NC1, NC2, NC3</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>NS2</td>
<td>NC4, NC5, NC6</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>
Transfer tests. The same measures (trials to criterion, accuracy, errors, and latency) were used to evaluate performance. Shifts in performance from the pretrained to novel color sets were assessed using the t test (related) and Wilcoxon test.

Cone excitation values of the colored stimuli. The differences between the cone excitation values for the chromatic and achromatic stimuli did not vary significantly from the normal distribution (C1–C7 = 0.245, df = 7, p = .2; C8–C15 = 0.161, df = 8, p = .2). The behavioral scores and the cone excitation values were compared using Pearson’s correlation coefficient for the parametric data and Spearman’s rank correlation coefficient for the nonparametric data.

The coefficient of determination ($r^2$) has been reported to indicate the proportion of explained variance as a percentage. This provides a way to express effect size as the proportion of variance in the variable to be predicted that can be explained by the predictor variable, which in this article are the measures of color discrimination performance and cone excitation values, respectively. Effects are reported as significant only if $p < .05$.

Results

Color Discrimination Training

All of the six horses successfully completed the color discrimination training, reaching the learning criterion of 10 consecutive correct choices ($p < .001$) for all 15 colors. Some variation in performance was found with individual colors. The color C7 (blue–green) took the greatest number of trials and C13 (orange) the least (see Figure 6). However, presenting the colors in the opposite order to the two groups controlled for any effect that order of training might have had, and these differences in relation to color were not found to be significant: trials to criterion, $\chi^2(14, N = 15) = 15.338, p = .355$; accuracy, $F(14, 70) = 1.085, p = .386$; errors, $F(14, 70) = 1.014, p = .45$; latency, $\chi^2(14, N = 15) = 2.832, p = .999$. The effect of color on performance accounted for 18.3% of the variation in the number of trials taken to reach the learning criterion, 17.8% of the accuracy, 16.9% of the number of repeated errors, and 3.4% of the variation in the speed of approach.

There was no significant difference between the groups trained with the different sequences of the colored stimuli for any of the measures used: trials to criterion, $U = 99, N_1 = N_2 = 15, p = .575$; accuracy, $t(28) = -0.289, p = .775$; repeated errors, $t(28) = 0.040, p = .968$; latency, $U = 101.5, N_1 = N_2 = 15, p = .648$. However, individual differences in the number of trials taken to reach the criterion are shown in Figure 7. It can be seen that, although performance with color C7 was consistently poor for Group 1 horses that had started their training at the short wavelength end of the spectrum (see Figure 7A), two of the horses in Group 2 reached the learning criterion in 10 and 12 trials, respec-
tively. Only one of the horses that had started the training at the long wavelength end of the spectrum had a problem reaching the learning criterion for this color (see Figure 7B).

Transfer Tests

The only significant difference in performance when transferring from the color discrimination training to the color set testing was in the speed at which the stimuli were approached \((z = -2.201, N - ties = 6, p = .028)\). Although on average the first color set was approached significantly more slowly than the final color trained, no drop in performance was found for any of the other measures: trials to criterion, \(z = -0.405, N - ties = 5, p = .686\); accuracy, \(t(5) = -0.643, p = .548\); errors, \(t(5) = 1.002, p = .362\).

However, it took a significantly greater number of trials to reach the criterion for this first set than for the second set, \(z = -2.023, N - ties = 5, p = .043\). No significant difference was found between any of the other consecutively tested sets, regardless of their colors. See Figure 8 for the mean number of trials taken to reach the criterion for the last color trained and each of the color sets. No significant overall difference was found in the accuracy percentages for the seven sets of colors; the improvement shown in Set 2 as compared with Set 1 was only marginal, \(t(5) = -2.323, p = .068\). The number of repeated errors made varied according to the set of colors being tested, \(F(6, 30) = 2.631, p = .036\), but when consecutive sets were compared, the only significant difference found was again between Sets 1 and 2, \(t(5) = 3.280, p = .022\), with fewer errors being made in the latter. No significant differences in the speed of approach were found with any of the sets of colors.

The introduction of the novel sets of colors was not found to have any effect on the measures taken to assess performance. It was immaterial as to whether the novel set introduced first consisted of a brown, green, and yellow stimulus for the first group, or a purple, blue, and pink stimulus for the second group: trials to criterion, \(z = -0.674, p = .51;\) accuracy, \(t(4) = 0.116, p = .48\); errors, \(t(4) = -0.377, p = .725\); latency, \(z = -1.091, p = .275\). Once the horses had learned to select the chromatic from the achromatic stimulus, regardless of the specific color, they were able to apply this rule to all of the pretrained colors equally easily, as well as to the two sets of novel colors.

Correlation With Cone Excitation Values

It was found that those colors that differed most from the small to medium-large cone excitation ratio of the achromatic stimuli (and as such were predicted as being the most colorful to the horse) were also most easily discriminated from grays. Cone excitation values were most predictive of accuracy scores for colors C1–C7 in that there was a highly significant positive correlation \((r = .923, n = 7, p = .003)\) between these measures, and \(r^2\) showed that the proportion of the variance explained was 85.2%. A negative correlation was found between the extent to which the cone excitation ratios differed from gray and both the number of repeated errors made \((r = -.880, n = 7, p = .009)\) and the latency to approach the stimuli \((r = -.775, n = 7, p = .041)\). The effect sizes based on \(r^2\) showed that the proportion of the variance explained was 77.4% and 60.1%, respectively. The correlation between the cone excitation ratio difference and trials to criterion was not significant \((p = -.679, n = 7, p = .094)\) and, as would be expected, the effect size was smaller (46.1%).

The analysis of colors C8–C15 (presented separately from the analysis of colors C1–C7 because of possible differences in the contribution to the perception of brightness made by each of the two cone types; Stockman & Sharpe, 1999) showed a significant negative correlation between the cone excitation ratio difference and the number of trials to criterion \((p = -.857, n = 8, p = .007)\). The greater the difference in the cone excitation ratio from the grays, the greater the accuracy \((r = .872, n = 8, p = .005)\), the fewer the repeated errors \((r = -.883, n = 8, p = .004)\), and the faster the selection was made \((p = -.792, n = 8, p = .019)\). Effect sizes (based on \(r^2\)) showed that the proportion of the variance explained was 73.4% for trials to criterion, 76.0% for accuracy, 78.0% for repeated errors, and 62.7% for latency to approach.

The relationship between the cone excitation ratio differences and the mean number of trials taken to reach the learning criterion for the 15 training colors is shown in Figure 9.

Discussion

All of the horses in the study demonstrated the ability to use chromatic information to discriminate between stimuli across the spectrum. They successfully reached the learning criterion for all of the colors, including five shades of green. As was previously found in the study by Grzimek (1952), horses can discriminate greens from grays when the stimuli are presented at ground level, as can the fallow deer (Birgersson et al., 2001). Like the fallow deer, the horse is a selective grazer and the ability to use the color of plants, as well as olfactory, gustatory, and tactile signals, as an additional cue in their choice of food may be an advantage. The recognition of familiar conspecifics, of prime importance in a herd animal such as the horse, would also be enhanced.

It was clearly demonstrated in the current study that the horses were using chromatic cues to make their selection by the way in which they applied this rule of “color is correct” in the transfer tests. Generalization to novel stimuli occurred equally well when the set contained a new green, brown, and yellow as when it contained a new blue, pink, and purple. Neither combination resulted in a decrease in discriminative ability. It is also worth noting that the inclusion of the color that had caused the most problems in the initial color discrimination training (C7) in Set 5
did not result in any decrease in performance either. Once the colors had been trained, they appeared equally easy for the horses to discriminate from grays.

However, in agreement with previous studies, although all the middle wavelength colors (greens) were successfully discriminated from grays eventually, more trials were required to reach the learning criterion for one color. Performance with C7 (blue–green) was less consistent, and more training was required (before the criterion was reached) than for any of the other colors. The study by Pick et al. (1994) resulted in the single subject failing to reach the same criterion (of 10 consecutive correct choices) when discriminating green from gray. Despite the use of an alternative criterion of 70% correct responses, Macuda and Timney (1999) also concluded that chromatic discrimination was inconsistent with green and yellow. It is likely that the green, C7, was close to the neutral point of the horse and appeared less colorful to the horse than the other shades of green. However, the reflectance spectra of the colors in the current study consisted of a broad range of wavelengths, as is the case with most natural objects, so some chromatic information was still available to the horse. The dominant wavelength (496 nm) of the green that could not be reliably discriminated from gray in the study by Pick et al. (1994) was also close to the neutral point location (480 nm) identified in the horse (Geisbauer et al., 2004), although that used by Macuda and Timney (1999) was not (530 nm).

It is possible that during the initial color discrimination training the horses had learned to avoid the negative stimulus (the gray stimulus cards) rather than actively selecting the positive one (containing the colored panel). However, novel grays were used in the transfer tests and caused no decrement in performance. The only cue to selection was that one pattern contained a chromatic panel; neither the grays nor the color presented was familiar to the horses. A further possible constraint on interpretation arises in that the panels that appeared gray to the human observer were not necessarily gray to the horse. Thus, in principle, in both the initial training and in the transfer tests the horses could have been basing their selection on the fact that one stimulus card contained a panel of a different color to that of the surrounding panels. However, gray card experiments have been used extensively in color vision studies in other species, and this phenomenon has not been reported previously (Kelber et al., 2003). Thus, this would seem to be an implausible explanation of the current results.

In addition to the possible effect of stimulus position on the results of earlier studies, discrepancies in previous findings may also be partially explained by individual differences in the horses used. Most features of visual ability are unlikely to be unique to individual animals, and the results gained from a few can be applied to the species as a whole. All of the horses in the current study successfully learned to discriminate all of the colors presented, but some variation between individuals was noted and the effect of previous experience was clear. Individual variation was also found in the five horses tested by Smith and Goldman (1999): Three horses successfully reached the learning criterion of 85% correct responses with green and yellow, but another performed at chance levels only with those colors. In an early study of color vision in the horse, Grzimek (1952) used only two subjects, as did Macuda and Timney (1999). Pick et al. (1994) trained a single horse of 19 years of age. Just as variation in color vision has been identified in normal human trichromats (Dartnall, Bowmaker, & Mollon, 1983), such individual differences may also occur in horses.

Although no statistically significant overall difference was found between the two groups of horses, their performance with one color (C7) was particularly variable. The horses that had previously reached the learning criterion with the medium-long wavelength colors (red, yellow, and green) had less of a problem attaining the criterion for C7 than did those that had previously been trained with the short wavelength colors (purples and blues). It is likely that the medium-long wavelength colors (including C7) appeared similar to the horse but different from the short wavelength colors. This finding is consistent with how the visible spectrum would appear to a dichromat: two hues divided by a neutral area (Sharpe, Stockman, Jägle, & Nathans, 1999).

The correlation found between the cone excitation values for each of the colors and the results of the color discrimination training was stronger for colors C8–C15 (the medium-long wavelength colors that would appear green, yellow, or red to the trichromatic human) than for C1–C7 (the short wavelength colors that would appear blue or purple). For example, yellow (C12) and orange (C13) were the two colors that varied the most from the achromatic stimuli in the proportion to which they excited the M–L cone (545 nm) photopigment. The latter color was the easiest to discriminate overall.

One of the consequences of dichromacy in the human is the inability to discriminate between certain colors, most commonly between certain shades of red and green (Jameson & Hurvich, 1978). The cone excitation data would provide a way of predicting likely difficulties with pairwise discriminations. The proximity of C3, C4, and C5 suggests that blues may be difficult to discriminate from purple. Given that C10 and C13 vary to the same extent from the achromatic stimuli, the horse may also have difficulty discriminating between lime green and orange.
Conclusions and Implications

The color discrimination training and testing provided a clear demonstration of horses’ ability to use chromatic information across the spectrum, including those wavelengths perceived as greens and yellows. The high correlation between the behavioral data and the calculated effects of the colors used on the cone photopigments of the horse suggests that cone excitation calculations could be used to predict the ease with which colors can be discriminated by the horse, as well as providing further evidence of dichromacy in this species. Such predictions would also provide some basis for the comparison of horse color perception with that of the trichromatic human and, thus, an improved understanding of equine behavior.

References


Received April 3, 2006
Revision received August 3, 2006
Accepted August 17, 2006

448 HALL, CASSADAY, VINCENT, AND DERRINGTON