

Color generalization across hue and saturation in chicks described by a simple (Bayesian) model

School of Experimental Psychology, University of Bristol,
Bristol, United Kingdom

Christine Scholtyssek

School of Life Sciences, University of Sussex, Brighton,
United Kingdom



Daniel C. Osorio

School of Life Sciences, University of Sussex, Brighton,
United Kingdom



Roland J. Baddeley

School of Experimental Psychology, University of Bristol,
Bristol, United Kingdom



Color conveys important information for birds in tasks such as foraging and mate choice, but in the natural world color signals can vary substantially, so birds may benefit from generalizing responses to perceptually discriminable colors. Studying color generalization is therefore a way to understand how birds take account of suprathreshold stimulus variations in decision making. Former studies on color generalization have focused on hue variation, but natural colors often vary in saturation, which could be an additional, independent source of information. We combine behavioral experiments and statistical modeling to investigate whether color generalization by poultry chicks depends on the chromatic dimension in which colors vary. Chicks were trained to discriminate colors separated by equal distances on a hue or a saturation dimension, in a receptor-based color space. Generalization tests then compared the birds' responses to familiar and novel colors lying on the same chromatic dimension. To characterize generalization we introduce a Bayesian model that extracts a threshold color distance beyond which chicks treat novel colors as significantly different from the rewarded training color. These thresholds were the same for generalization along the hue and saturation dimensions, demonstrating that responses to novel colors depend on similarity and expected variation of color signals but are independent of the chromatic dimension.

Introduction

Birds have four spectral types of narrowly tuned single cones allowing them to discriminate a huge

variety of colors. The ability to discriminate small color differences is important for some behaviors, such as judging potential mates (Cuthill, Bennett, Partridge, & Maier, 1999; Fitzpatrick, 1998), but fine distinctions on the color continuum may be undesirable in other contexts, as when food items are edible or inedible over a range of discriminable colors. It would not pay to learn separately about the inedibility of every possible color of ladybirds—and the fact that warning colors of ladybirds vary (Bezzarides, McGraw, Parker, & Hussein, 2007) suggests that this is not what birds do. Even the color of a single object can vary between encounters, for instance when viewed on different backgrounds or when color constancy fails (Osorio, 2009). Hence, to understand how birds use color to recognize objects, it is important to study how they generalize across discriminable, yet similar colors.

Most studies of how birds generalize color have used monochromatic stimuli (see, e.g., Ghirlanda & Enquist, 2003, and references therein). Birds, often pigeons, are trained to respond to a certain wavelength, say 550 nm, and their responses to novel wavelengths, say 555 and 560 nm, are then compared to the trained wavelength to obtain a generalization gradient. These gradients can be described by either a Gaussian or an exponential function of the distance from the training wavelength (Ghirlanda & Enquist, 2003). A difficulty in studying color generalization on a wavelength continuum is that it disregards the birds' ability to discriminate between the different wavelengths in the test; the discrimination threshold varies substantially as a function of wavelength

Citation: Scholtyssek, C., Osorio, D. C., & Baddeley, R. J. (2016). Color generalization across hue and saturation in chicks described by a simple (Bayesian) model. *Journal of Vision*, 16(10):8, 1–10, doi:10.1167/16.10.8.

doi: 10.1167/16.10.8

Received February 18, 2016; published August 18, 2016

ISSN 1534-7362



(Emmerton & Delius, 1980), which distorts the shape of the generalization gradients so that most studies cannot distinguish responses to novel colors due to perceptual confusion from those due to sensory generalization. This study accounts for sensory confusion by specifying the discriminable distance of novel colors from the training colors by an empirically verified model of color discrimination (Vorobyev & Osorio, 1998).

Monochromatic lights, used in most former studies of color generalization, cannot excite more than two single cone types at once in the bird eye, and therefore they vary almost exclusively in hue. By comparison, our experiments use printed stimuli, which reflect light like natural surfaces, and their broad reflectance spectra can excite multiple cone types, allowing the color to vary along an additional chromatic dimension: saturation, which is the difference of a color from an achromatic gray (Wyszecki & Stiles, 1982).

Hue and saturation are familiar to humans as distinct aspects or dimensions of color, but their significance in animal perception and visual communication signals is little known. In general, one might expect that while hue will depend on the chemical composition of the pigment, saturation of a color will depend on both the pigment concentration and possibly any surface covering, such as dust. Great tits (*Parus major*) vary in the level of yellow pigment of their breast feathers, and it has been suggested that hue signals individual foraging success and saturation overall body condition, providing independent signals during mate choice (Senar, Negro, Quesada, Ruiz, & Garrido, 2008). More generally, the light reflected from a uniform specular surface (such as a feather or beetle elytron) will vary in saturation but not hue, with the chromaticity of points on the surface lying on a line between the color locus of the illumination (i.e., achromatic) and the locus of the material seen with minimum specular reflectance. Consequently, hue and saturation may give different types of information about objects and surfaces and hence have different behavioral significance.

If birds differentiate hue from saturation as independent sources of information in decision making, do they react differently to suprathreshold color differences in these two aspects of color? Birds encountering novel colors could innately weight changes in one dimension more strongly than changes in the other dimension, resulting in different degrees of generalization along the different chromatic dimensions. Alternatively, birds might weight changes in either dimension according to their experience of how the signals vary. If birds learned that the same magnitude of color change signals equal changes in profitability (e.g., mate quality or nutritional value of

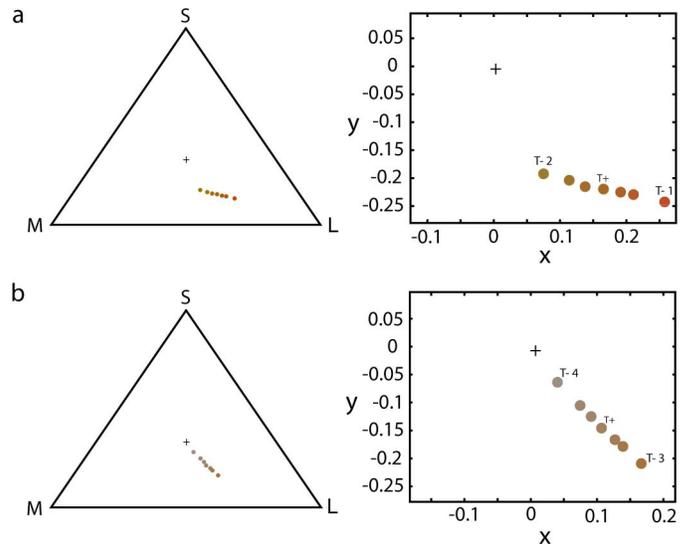


Figure 1. Training and testing colors represented in the chicks' receptor space. Vertices of the Maxwellian triangle correspond to the excitation of the chicks' L, M, and S cones. (a, b) Training and testing stimuli along opposite directions in the hue and saturation dimensions, respectively. T– 1–4 label unrewarded training stimulus for Groups 1–4; T+ labels the rewarded training stimulus. The symbol + marks the location of the achromatic background.

food), then generalization of hue and saturation should be the same. To our knowledge, generalization gradients have not been described along the saturation dimension in birds or in any other animal (including humans). Here, to directly compare how poultry chicks generalize to novel colors along either dimension, we use behavioral experiments and statistical modeling of generalization of hue and saturation. Chicks are excellent subjects for tests of color generalization: They learn colors fast and accurately, and from hatching we can control their color experience and hence knowledge about color variation. Also, because the spectral sensitivities of chicken cone photoreceptors and color discrimination thresholds are known, we can produce colors of known discriminability from one another to test generalization.

In our experiments chicks learned to forage on colored food containers and to identify colors signaling the presence or absence of a reward. The rewarded and unrewarded colors were thereby separated by the same suprathreshold color distance on either continuum, dominated either by hue or by saturation changes in the chicks' receptor space (Figure 1).

In tests we simultaneously presented food containers printed in the training colors and novel intermediate colors, and we recorded the chicks' choice preferences under extinction (i.e., in the absence of a food reward). We hypothesize, in accordance with the

Distance from T+	R	G	B	x	y	Log brightness
Hue						
4.03	0.604	0.484	0	0.114	−0.206	4.45
2.19	0.631	0.451	0	0.075	−0.193	4.45
1.15	0.649	0.429	0	0.138	−0.215	4.47
0	0.672	0.402	0	0.165	−0.219	4.47
1.11	0.694	0.374	0	0.192	−0.225	4.47
1.98	0.717	0.347	0	0.211	−0.230	4.44
4.03	0.766	0.286	0	0.258	−0.243	4.45
Saturation						
4.11	0.617	0.558	0.5	0.04	−0.064	4.45
2.14	0.627	0.532	0.436	0.075	−0.105	4.44
0.98	0.633	0.519	0.405	0.091	−0.125	4.44
0	0.637	0.508	0.378	0.107	−0.146	4.45
1.05	0.654	0.489	0.333	0.127	−0.167	4.45
2.02	0.651	0.473	0.295	0.139	−0.179	4.46
4.11	0.662	0.445	0.222	0.167	−0.209	4.49
Mean background	0.6	0.6	0.6	0	0	4.46

Table 1. Properties of the training and testing stimuli and the achromatic background. *Notes:* From left to right, the table shows the discriminable distance (in just-noticeable differences) of each color from T+, the RGB values of the lookup table used to generate the colors in MATLAB (from 0 to 1), x- and y-coordinates in the chicks' receptor space (Figure 1), and log brightness (relative quantum catches) as calculated for the chicks' double cones.

matching law (Herrnstein, 1970), that the relative frequency with which the chicks choose the different testing colors is proportional to their estimation of the probability of a reward. To characterize how chicks generalize the association of a reward from the training to the testing colors, we introduce a Bayesian model that converts the chicks' choice behavior to a probability function of the distance of each testing color from the rewarded training color. To compare generalization between the different training and testing conditions, we propose a summary statistic of generalization behavior, which is not provided by former studies on sensory generalization. This Bayesian model allows us to interpolate a generalization threshold from our fitted probability function and to compute the confidence in this measure. The threshold describes a color distance beyond which the chicks treat novel colors as significantly different from the rewarded training color. This new method for calculating psychophysical discrimination thresholds has advantages over standard psychophysical methods of generating a psychometric function, which require the successive comparison of the rewarded standard stimulus to a variety of comparison stimuli in a large number of separate trials (Gescheider, 1976). In contrast, our Bayesian method of computing the threshold allows the simultaneous comparison of an arbitrarily large number of physical or perceptual alternatives to generate the psychometric or generalization function. At the cost of an assumption of a constant threshold in any given dimension, this method requires an order of magnitude less data than

the standard method. Furthermore, the Bayesian method presented here can return a threshold that is beyond the range tested. Therefore, our model provides not only a long-overdue approach to directly comparing generalization performance within or between stimulus dimensions but also a useful opportunity to study sensory discrimination.

Methods

Stimulus design

Stimulus patterns were printed on standard plain paper with a Canon PIXMA Pro9500 ink-jet printer and folded into conical food containers (Osorio, Vorobyev, & Jones, 1999). The patterns consisted of 48 tiles (each 6×2 mm); 15 randomly assigned tiles were printed in the stimulus color, and the remaining 33 were printed in gray. To eliminate luminance cues, colors were chosen to be isoluminant for the chicks' double cones (Table 1; Supplementary Figure S1). In addition, the luminance of the 33 gray tiles was varied with a Michelson contrast of 30%. Two shades of gray were identified that were 15% brighter and darker than the stimulus color. Between these two gray values, 33 homogeneous gray intervals were calculated and randomly assigned to the 33 achromatic tiles. This way, the average luminance of all gray tiles matched the luminance of the colored tiles.

The stimuli were illuminated by a quartz halogen light source, which was long-pass filtered (Schott GG475, Schott, Mainz, Germany) to remove most of the spectrum that stimulates the VS cones. As can be seen from Supplementary Figure S1, the amount of light that could stimulate the VS cones is small, and given the low numbers of VS cones (e.g., Bowmaker, Heath, Wilkie, & Hunt, 1997), quantum catch is very low, so that VS cones are most unlikely to have made a contribution to color discrimination. Therefore, the chicks discriminated colors with the remaining three types of single cones (S, M, L; Osorio et al., 1999).

The colors were chosen to fall on straight lines in the chicks' receptor space (Figure 1). Colors varied either exclusively in saturation or simultaneously in hue and saturation. This is due to the use of isoluminant printed stimuli. "Yellower" colors at similar saturations are brighter for the chick double cones than "redder" colors. To exclude brightness cues, we had to adjust the luminance of the printed colors, which made colors less saturated as they contained more yellow. Color measurements and modeling of the color coordinates are described in Supplementary Figure S1. The perceptual distance between colors was calculated using the receptor-noise limited model of Vorobyev and Osorio (1998; see also Table 1; Supplementary Figure S1), which has recently been validated experimentally (Olsson, Lind, & Kelber, 2015). Furthermore, it is reasonable to assume that equally distant stimuli in color space are also equally discriminable (unlike monochromatic stimuli, where discriminability will systematically vary with wavelength).

Procedure

Five- to 10-day-old untrained male poultry chicks were trained to discriminate between paper cones of a rewarded and an unrewarded color separated by four just-noticeable differences (JNDs) on either a hue or a saturation continuum (Figure 1; Table 1; Supplementary Figure S1). Twenty-four pairs of chicks were assigned to four different training groups. Groups 1 and 2 received differential training along the hue continuum. Both groups were trained on the same rewarded color (T+), which appeared orange to human observers, but the unrewarded color (T−) was located at opposite ends of the hue continuum, with Group 1 being trained with a "redder" T−, and Group 2 with a "yellower" T− (Figure 1; Table 1). Groups 3 and 4 received differential training along the saturation dimension (Figure 1; Table 1). Here as well, both groups were trained on the same T+, which was an intermediate saturated orange, but the T− was either more saturated (Group 3) or less saturated (Group 4) than the T+.

Chicks were trained in pairs to forage on four T+ and four T− containers, which were randomly dispersed over the floor of a 30 × 30 cm arena. T+ containers were filled with chicken crumbs, while T− containers remained empty. The chicks rapidly learned to extract food from the containers by pecking on them. T+ containers were refilled at 1-min intervals, six times during a training session. Two sessions were run per day, for a total of three training days.

Two tests were run on two subsequent days, with one additional training session preceding each test. Before training and testing, chicks were deprived of food for 2 hr. Tests were carried out in extinction (i.e., without reward). Chicks were presented with the T+ color, the T− color, and two intermediate colors (Figure 1), which were separated from the T+ by approximately one and two JNDs (Table 1). Each color was present twice. In each test, choice frequencies for the four colors were recorded for one chick in each pair, until eight choices were completed. The number of responses to each stimulus were pooled for the six chicks within a group and plotted as a function of the discriminable distance of the test stimuli from T+.

All experiments were conducted in line with the United Kingdom's Animals (Scientific Procedures) Act.

The Bayesian model

Our data consist of the number of choices made by the chicks from N alternatives, each characterized by its discriminable distance from the T+ color. To compare generalization between conditions, it is useful to translate our choice data into a model of the chicks' certainty about a reward and how that certainty depends on the perceptual distance of novel colors from the T+ color. We assume that the chicks peck on the N alternative stimuli with a probability proportional to the probability that the stimulus is rewarded (Herrnstein, 1970). In accordance with the ideal observer theory (e.g., Geisler, 2003), the probability that a chick will choose a stimulus k with a discriminable distance X_k to the rewarded training stimulus T+ can be determined by invoking Bayes's rule:

$$P(\text{choice} = k | X_k) = \frac{P(X_k | \text{choice} = k)P(\text{choice} = k)}{\sum_j P(X_j | \text{choice} = j)P(\text{choice} = j)}. \quad (1)$$

The expression $P(\text{choice} = k)$ is our belief about the probability of a chick's choosing stimulus k prior to any training. If we assume no pretraining bias such as innate color preferences, $P(\text{choice} = k)$ is the same for all N stimuli, and is therefore simply $1/N$. The expression $P(X_k | \text{choice} = k)$ is the probability of X_k given the choice of stimulus k . This term is called the likelihood function. Since the prior probability

$P(\text{choice} = k)$ is constant, the likelihood function is the only term we have to fit to describe the probability of a chick's choosing stimulus k given its discriminability X_k from T+ (Equation 1). The denominator is the total probability of X_j for all stimuli.

We compare the fits of two likelihood functions to the data. First, given that decision making is likely to be influenced by a number of factors, including neural noise, speed-versus-accuracy trade-off, and the estimate of the costs and benefits of the choice, the central limit theorem predicts that the likelihood function is Gaussian, so that

$$P(X_k | \text{choice} = k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X_k - T)^2}{2\sigma^2}\right). \quad (2)$$

$(X_k - T)$ measures the number of JNDs a given stimulus is away from the rewarded pattern. This means that the standard deviation σ is the only parameter that needs to be fitted.

We also explore a Laplace likelihood function (a likelihood that is an exponentially decaying function from the reference rather than a Gaussian), as suggested by Shepard (1987) and Tenenbaum and Griffiths (2001):

$$P(X_k | \text{choice} = k) = \frac{1}{2a} \exp\left(\frac{-|X_k - T|}{a}\right). \quad (3)$$

Exponential likelihood functions can plausibly be motivated by in fact having Gaussian likelihoods but with an unknown standard deviation. Averaging (marginalizing) over this uncertainty can in some situations result in a so-called Laplace likelihood. Like the standard deviation σ in Equation 2, a determines the width of the function and is the only parameter to fit.

Fit

To estimate parameters and compute the confidence in the estimate, we use established methods, as for instance described in the reviews of Wichmann and Hill (2001a, 2001b). To obtain a point estimate of the parameter for the Laplace or the Gaussian fit, the model uses maximum log likelihood estimation recruiting a MATLAB (The MathWorks, Natick, MA) built-in unconstrained nonlinear minimization search algorithm (fminsearch.m).

To decide which type of function best describes our data, the model performs nonparametric bootstraps. For $n = 1,000$ iterations, new data sets are created by randomly sampling from the original data set with a sample size equivalent to the original data. For each new data set the model performs a maximum (log) likelihood estimation for both types of fit and compares them in a likelihood ratio test. Since we compute log likelihoods,

we can calculate the difference instead of the ratio:

$$r = \log L(\sigma | D') - \log L(a | D'), \quad (4)$$

with $L(a | D')$ being the likelihood given the Laplace function and $L(\sigma | D')$ being the likelihood given the Gaussian function for a given data set D' . This way we obtain a distribution of n likelihood ratios r , of which the mean and the 95% confidence limits are calculated (explicitly, the values of the 97.5 and 2.5 percentiles). This nonparametric bootstrap is performed by the function `fit_exp_and_gauss.m` in MATLAB (available on request), which returns a distribution of bootstrapped log likelihood ratios.

After deciding which type of fit describes the data best, the model computes the confidence in the fit by sampling from the posterior distribution using a Metropolis–Hastings algorithm (see, e.g., Wichmann & Hill, 2001b). After an initial burn-in period, for n iterations this algorithm approximates the posterior distribution of the parameter values by a set of samples, as is standard Bayesian modeling practice.

For simplification, from now on the symbol b will be used for the parameter, standing for either a or σ depending on the outcome of the likelihood ratio tests.

To minimize burn-in time, the Metropolis–Hastings algorithm uses the initial estimate of the value of b that maximizes the likelihood, as the initial value b_t . It proposes a candidate value b' that is randomly sampled from a normal distribution $P(b' | b_t)$ centered around b_t , and then compares the posterior (which here is the same as the likelihood, since we assume a uniform prior) of the current and the proposed model using the following standard Metropolis–Hastings acceptance ratio test. The posterior ratio (likelihood ratio, since we have equal priors) is calculated as

$$r = \frac{L(b' | D)}{L(b_t | D)}. \quad (5)$$

Here $L(b' | D)$ and $L(b_t | D)$ are the posterior probabilities (likelihoods) of the proposed and the current model, respectively, given our choice-frequency data D .

If the proposed model is more likely than the current model ($r > 1$), b' is chosen as a sample of the posterior probability distribution of b . If the proposed model is less likely than the current model ($r < 1$), b' is rejected with a probability equivalent to r and b_t is chosen instead. The chosen value will then serve as the current value b_t in the next iteration. Repeating this process allows us to obtain samples from the posterior probability distribution together with its 97.5 and 2.5 percentiles (serving as a measure of our 95% confidence interval).

To estimate how well the model describes the data, we calculate the respective correlation coefficient R^2 that describes the percentage of variance in the data that the model accounts for.

Threshold interpolation

To compare the generalization among the four groups, we compare difference thresholds: a color distance beyond which novel colors are treated significantly different from T+, or in other words a distance at which the chick discriminates between T+ and novel colors. We adopt the threshold criterion from standard psychophysical methods to determine sensory discrimination thresholds. Both types of thresholds are determined by uncertainty. Whereas in sensory discrimination, performance is limited by actual noise in the sensory neurons or by stimulus-inherent variations in appearance, in the present study performance is limited by uncertainty about the presence of a reward (Lynn, 2010). In the case of sensory discrimination, the threshold describes the minimal physical distance necessary to yield a just-noticeable (perceptual) difference. In terms of discrimination thresholds, “just noticeable” was defined as the difference that is detected some proportion of the time. Likewise for generalization, this means that the stimulus is treated differently (is discriminated from T+) some proportion of the time. In psychophysics, this proportion, known as the threshold criterion, is usually a performance halfway between chance level and 100% T+ choices, and is used to interpolate the threshold from a model fitted to the data. In a two-alternative forced-choice task in which the subject is asked to choose the T+ over a simultaneously presented comparison stimulus, the chance level is 50%. Therefore the threshold would be a distance from T+ at which T+ is chosen in 75% of all presentations. In a simultaneous four-alternative forced-choice task, the chance level would be 25%, so the threshold criterion would be 62.5%. However, this is only true if the three comparison stimuli are *identical*. Here, we compare four *different* colors simultaneously, which all have a different choice probability that has to be considered (Equation 1). Since our model converts choice frequencies into probabilities we can apply a two-alternative forced-choice threshold criterion, despite using a multiple-choice paradigm to generate the data by asking for the relative probability $P_r(\text{T+})$ of T+ compared to any novel color with a measurement value X :

$$P_r(\text{T+}) = \frac{P(\text{T+})}{P(\text{T+}) + P(X)}, \quad (6)$$

with $P(\text{T+})$ as the probability of T+ and $P(X)$ as the probability of a color at a discriminable distance X . For colors very close to T+, $P_r(\text{T+})$ is around 50% and increases with increasing discriminability of novel colors from T+ until it reaches 100% for colors that are very different from T+. Therefore, we can interpolate the difference threshold using a criterion of $P_r(\text{T+}) = 75\%$. If we assume a Gaussian or a Laplace likelihood function (Equations 2 and 3), and disregard normalization,

Equation 6 can be written as

$$P_r(\text{T+}) = \frac{1}{1 + \exp\left(-X^2/(2\sigma^2)\right)} \quad (7)$$

for a Gaussian likelihood function and

$$P_r(\text{T+}) = \frac{1}{1 + \exp(-|X|/a)} \quad (8)$$

for a Laplace likelihood function. As becomes obvious from Equations 7 and 8, the thresholds can be directly calculated from the estimated parameters (σ or a). Assuming a threshold criterion of 0.75 (75%), solving for X reveals a linear relationship between the parameter and the threshold, with $X_{\text{threshold}} = 1.482\sigma$ for a Gaussian likelihood function and $X_{\text{threshold}} = 1.099a$ for a Laplace likelihood function. This threshold interpolation is performed for each parameter value in the posterior distribution, which was obtained using the Metropolis–Hastings algorithm, as described in the previous section (Equation 5). This way a probability distribution of thresholds with mean and upper and lower 95% confidence limits of the mean can be obtained as summary statistics and a measure of confidence to compare generalization between conditions.

The function (`fit_softmax.m`) that fits the model and interpolates the threshold was programmed in MATLAB. The only input `fit_softmax.m` needs is a vector containing the distances of the test stimuli from T+, a vector containing the corresponding choice frequencies, and the threshold criterion. The output of the function is the mean threshold, the standard error of the thresholds, the 95% confidence interval of the threshold, and a figure illustrating the relative choice frequencies, the fitted model (Figure 3), and, optionally, the distribution of the thresholds.

For the illustration, the function computes the likelihood function for the entire posterior distribution of the parameter b obtained using the Metropolis–Hastings algorithm (see Fit) and plots the mean as well as the standard deviation as a function of the perceptual distance X from T+ (Figure 3).

Results

Four groups of chicks were trained to discriminate colors separated by about four JNDs on a continuum dominated by changes in either hue (Groups 1 and 2) or saturation (Groups 3 and 4). We then tested the chicks' generalization performance in extinction on each continuum by recording their choice frequencies for the T+, the T−, and two novel intermediate colors separated by roughly one and two JNDs from the T+ (Figure 1; Table 1). Each condition comprised a total of

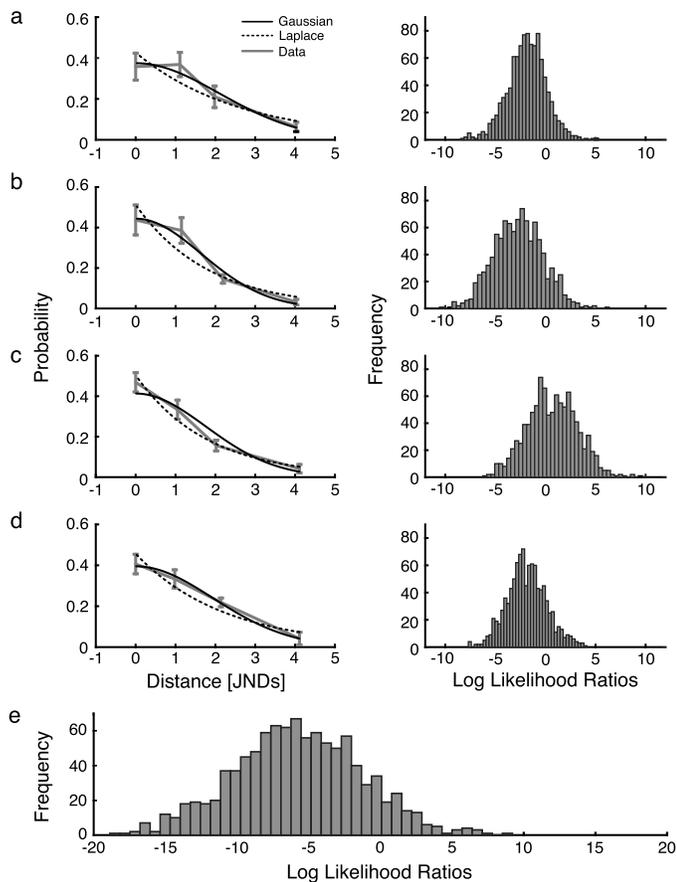


Figure 2. Comparison of a Laplace and a Gaussian fit using maximum likelihood estimation. The solid gray line in the left panels shows the relative choice frequencies for the four testing stimuli. Error bars depict standard errors for $n = 6$ chicks. The solid black line represents the Gaussian fit, and the dotted line the Laplace fit. The right panels depict histograms of the log likelihood ratios computed by means of nonparametric bootstrapping. Negative values prefer the Gaussian fit, positive values the Laplace fit. (a) Group 1, trained with an “orange” T+ and a “redder” T-. (b) Group 2, trained with the same T+ but a “yellower” T-. (c) Group 3, trained with more saturated T-. (d) Group 4, trained with a less saturated T-. (e) Distribution of the log likelihood ratios for all four groups.

96 choices (six chicks each making 16 choices in two tests). To compare performance along the hue and saturation dimensions, we employed a Bayesian model that converts choice frequency into a posterior distribution of the discriminable distance of novel colors from T+. We explored a Gaussian and a Laplace function to fit our choice-frequency data. To estimate which of the two functions best describes the color-generalization data, we performed a nonparametric bootstrap and compared the maximum likelihoods for the Gaussian and the Laplace function in likelihood ratio tests (Equation 4).

A comparison of the Laplace and the Gaussian fits and the distributions of the log likelihood ratios for

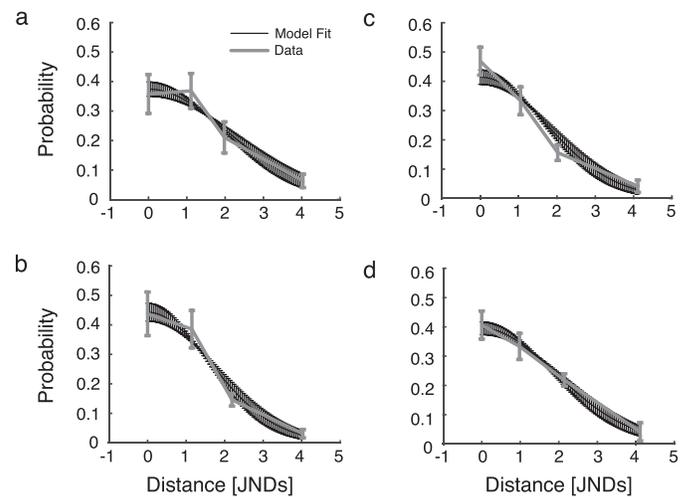


Figure 3. Generalization functions and threshold distributions for chicks tested along opposite directions on the hue and saturation dimensions. The solid gray line in the left panels shows the relative choice frequencies. Error bars depict standard errors for $n = 6$ chicks. The black curve shows the mean and the standard deviation of the Gaussian function fitted to our data. a–d represent fits for Group 1–4, respectively.

each of the four groups are shown in Figure 2a through d. If either of the functions were to provide a significantly better description of the data, the 95% confidence interval of the likelihood ratio should exclude 0 (at which both functions have the same likelihood given the data). However, this is not the case for any of the four groups (Figure 2a through d, right panels), showing that both functions are suitable for describing color generalization by chicks. Assuming the same likelihood for all four groups, the Gaussian function seems to fit better (yet not significantly so) than the Laplace function (Figure 2e), with a mean log likelihood ratio of -5.66 and confidence limits at -14.53 and 2.75 . Therefore, we decided to use this type of function to further explore the chicks’ generalization performance.

Figure 3 depicts the mean and standard deviation of the Gaussian fits obtained by the Metropolis–Hastings algorithm (Equation 5). For all four groups, the model provides a very good description of the chicks’ generalization performance, accounting for 93%–100% of the variance in the data (R^2 values of the mean fit for Groups 1–4 are, respectively, 0.97, 0.95, 0.93, and 1).

The mean and the upper and lower 95% confidence limits of σ obtained using parametric bootstrapping for Groups 1–4 are, respectively, 2.13 [1.73, 2.78], 1.67 [1.36, 2.12], 1.76 [1.43, 2.17], and 1.96 [1.59, 2.43].

As a measure of the degree to which chicks generalize the training behavior across novel colors, we interpolated generalization thresholds from the fitted model. The distributions of the thresholds are illustrated in Figure 4 as violin plots with mean and 95%

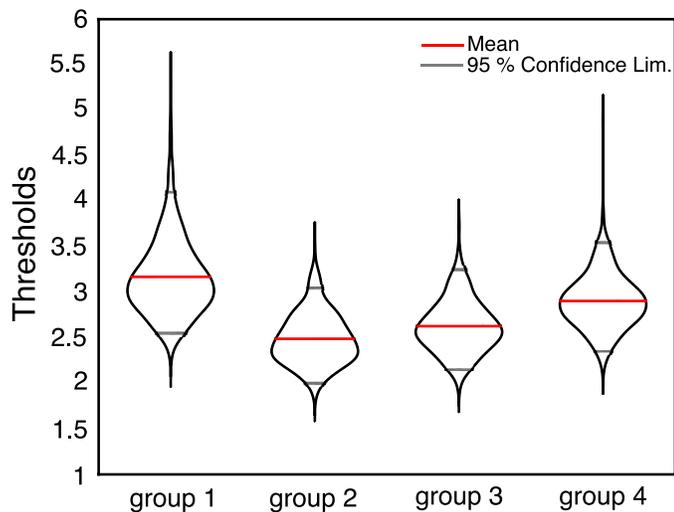


Figure 4. Violin plots of the thresholds. The violins are rotated and mirrored kernel-density estimations of the threshold distributions, illustrated with means and 95% confidence limits.

confidence limits. The mean and the upper and lower 95% confidence limits of the thresholds obtained for generalization along the two opposite directions of the hue continuum are 3.16 [2.55, 4.1] (Group 1) and 2.48 [2, 3.15] (Group 2). The mean and upper and lower 95% confidence limits of the thresholds obtained for generalization along the saturation continuum are 2.62 [2.15, 3.2] (Group 3) and 2.9 [2.35, 3.6] (Group 4). Although the mean generalization threshold of Group 1 is higher than that of Group 2, the 95% confidence intervals of the means between Groups 2, 3, and 4, as well as between Groups 1 and 4, strongly overlap, indicating a lack of any significant difference between generalization. Hence, generalization does not depend on the chromatic dimension in which training and testing colors varied. Furthermore, mean thresholds obtained for the two opposite testing directions on each dimension are remarkably similar, at 2.82 JNDs for hue and 2.76 JNDs for saturation.

Discussion

Modeling generalization and determining thresholds

We have introduced a Bayesian model that describes generalization as a probability or likelihood function of the discriminable distance between novel and known stimuli. Earlier analysis suggested that the shape of generalization gradients follows either a Gaussian or a Laplace function (Ghirlanda & Enquist, 2003; Shepard, 1987). Figure 2 shows that either distribution accounts very well for the chicks' generalization of color. This

might be due to the fact that we tested generalization over a small number of discriminable distances from T+. To discriminate between Laplace and Gaussian likelihood functions, denser sampling would be beneficial, especially close to T+. However, the minimum discriminable distance of our test stimuli is approximately one JND from T+, and sampling below that would reflect responses made due to perceptual confusion rather than sensory generalization. Hence, determining the exact shape of the generalization gradient appears to be difficult: The predictions of the two models are sufficiently similar that discriminating them would require considerably larger data sets than these.

As a measure of the degree of generalization we interpolated discriminable distances beyond which novel colors are treated as significantly different from T+. Our model thereby computes a posterior probability distribution of thresholds, which allows a measure of confidence in our threshold estimation. By comparing these posterior distributions of generalization thresholds, we found that color generalization in chicks does not depend on the chromatic dimension in which the colors vary. Instead, there was a minor difference in the distribution of the generalization thresholds for opposite directions of colors varying predominantly in the hue dimensions. This difference might result from an innate preference for red over yellow, as has been shown for imprinting by newly hatched chicks (Salzen, Lily, & McKeown, 1971). If chicks also have a red preference during foraging, this may increase the probability with which they choose novel training colors after being trained with a redder T− (Group 1). Conversely, the same preference may have decreased the probability with which the chicks choose novel colors following training with a yellower T− (Group 2). Interestingly, the threshold distributions of the two opposite directions in the saturation dimension were similar, although untrained young poultry chicks have an innate preference for more saturated oranges during foraging (Ham & Osorio, 2007).

The distinction between hue and saturation as different sources of information

In normal discourse and in color science, we conventionally distinguish hue and saturation (or color purity) as separate aspects of color, perhaps because changes in hue and saturation are likely to have different physical causes (see Introduction); but differences in perception of hue and saturation are little studied in animals. It has been suggested that hue and saturation in the yellow breast plumage of the great tit (*Parus major*) convey different types of information

about the bird's quality (Senar et al., 2008), and if hue and saturation can generally provide complementary information, then the ability to analyze these two chromatic dimensions of color independently when encountering novel colors would allow birds to optimize decisions—for example, about what to eat or which mate to choose. This raises the question of whether the chromatic dimension influences the degree to which birds generalize responses to novel colors.

Here we describe color-generalization gradients after training young domestic chicks to discriminate between rewarded and unrewarded colors, which are separated by the same suprathreshold color distance, for colors varying either mainly in hue or solely in saturation. Hence, chicks trained in either condition should have learned the same correlation between color variation and the presence or absence of a reward. We did not achieve isosaturation for the colors used to test generalization along the hue dimension. This is due to the problem of creating isoluminant colors. In fact, tests found no difference in the degree to which chicks generalize the association of a reward across novel colors, showing that generalization is not dependent on whether color differences are constituted by a change in saturation alone or by a simultaneous change in hue and saturation. Furthermore, tests found no bias toward more- or less-saturated colors. What seems to be important is perceptual similarity, defined by the number of JNDs separating known and novel colors, combined with knowledge about how colors vary. These findings are therefore consistent with Fechner's law, which predicts that the perceived magnitude of suprathreshold color differences should be proportional to their separation in JNDs (Ham & Osorio, 2007; Renoult, Kelber, & Schaefer, in press; Wyszecki & Stiles, 1982). Learning about variation in generalization behavior is likely to be adaptive, since it allows birds to constantly update their assumptions of the variation of object colors and therefore to adjust the probability with which they generalize behaviors to novel colors depending on their experience with previous colors. It may also allow them to generalize differently in different contexts. In the context of mate choice, for instance, small suprathreshold color changes may make all the difference between a high- and a low-quality mate, whereas in the context of foraging, small color differences may be negligible.

We have not tested here whether birds can learn that variation in one chromatic dimension is more or less important than variation in the other chromatic dimension, nor whether birds differentiate between hue and saturation as independent signals. To bring us one step closer to this understanding, we are currently training chicks to discriminate between two-dimensional color distributions that have defined variances in hue and saturation. The rewarded and unrewarded

color distributions are thereby separated by only one dimension, hue or saturation, making variation along the other dimension completely irrelevant. We will then test whether color generalization along the relevant dimension is influenced by the extent to which colors vary in the irrelevant dimension, and will be able to make inferences about the degree to which birds are able to disentangle the two chromatic dimensions (or any other possible dimensions) of color.

Keywords: color generalization, hue, saturation, Bayesian inference, multiple-choice paradigm, psychometric function

Acknowledgments

We thank Peter Olsson for helpful discussions on the calculations of receptor noise levels. We furthermore thank the two reviewers for valuable comments on an earlier version of the manuscript. This research was funded by a grant from the Biotechnology and Biological Sciences Research Council to RJB and DCO.

Commercial relationships: none.

Corresponding author: Christine Scholtyssek.

Email: cs508@sussex.ac.uk.

Address: School of Life Sciences, University of Sussex, Brighton, United Kingdom.

References

- Bezzaides, A. L., McGraw, K. J., Parker, R. S., & Husseini, J. (2007). Elytra color as a signal of chemical defense in the Asian ladybird beetle *Harmonia axyridis*. *Behavioral Ecology and Sociobiology*, *61*(9), 1401–1408, doi:10.1007/s00265-007-0371-9.
- Bowmaker, J. K., Heath, L. A., Wilkie, S. E., & Hunt, D. M. (1997). Visual pigments and oil droplets from six classes of photoreceptor in the retinas of birds. *Vision Research*, *37*(16), 2183–2194, doi:10.1016/S0042-6989(97)00026-6.
- Cuthill, I. C., Bennett, A. T. D., Partridge, J. C., & Maier, E. J. (1999). Plumage reflectance and the objective assessment of avian sexual dichromatism. *The American Naturalist*, *153*(2), 183–200, doi:10.1086/303160.
- Emmerton, J., & Delius, J. D. (1980). Wavelength discrimination in the visible and ultraviolet spectrum by pigeons. *Journal of Comparative Physiology*, *141*(1), 47–52, doi:10.1007/BF00611877.

- Fitzpatrick, S. (1998). Colour schemes for birds: Structural coloration and signals of quality in feathers. *Annales Zoologici Fennici*, 35(2), 67–77.
- Geisler, W. S. (2003). Ideal observer analysis. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 825–838). Cambridge, MA: MIT Press.
- Gescheider, G. A. (1976). *Psychophysics: Methods and theory*. New York: Lawrence Erlbaum Associates.
- Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behavior*, 66, 15–36, doi:10.1006/anbe.2003.2174.
- Ham, A. D., & Osorio, D. (2007). Colour preferences and colour vision in poultry chicks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621), 1941–1948, doi:10.1098/rspb.2007.0538.
- Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13(2), 243–266, doi:10.1901/jeab.1970.13-243.
- Lynn, S. K. (2010). Decision-making and learning: The peak shift behavioral response. In M. Breed & J. Moore (Eds.), *Encyclopedia of animal behavior* (Vol. 1, pp. 470–475). Cambridge, MA: Academic Press.
- Olsson, P., Lind, O., & Kelber, A. (2015). Bird colour vision: Behavioural thresholds reveal receptor noise. *Journal of Experimental Biology*, 218(2), 184–193, doi:10.1242/jeb.111187.
- Osorio, D. (2009). Color generalization in birds. In L. Tommasi, M. A. Peterson, & L. Nadel (Eds.), *Cognitive biology* (pp. 129–146). Cambridge, MA: MIT Press.
- Osorio, D., Vorobyev, M., & Jones, C. D. (1999). Colour vision of domestic chicks. *Journal of Experimental Biology*, 202(21), 2951–2959.
- Renoult, J. P., Kelber, A., & Schaefer, H. M. (in press). Colour spaces in ecology and evolutionary biology. *Biological Reviews*, doi:10.1111/brv.12230.
- Salzen, E. A., Lily, R. E., & Mckeown, J. R. (1971). Colour preference and imprinting in domestic chicks. *Animal Behaviour*, 19(3), 542–547, doi:10.1016/S0003-3472(71)80109-4.
- Senar, J. C., Negro, J. J., Quesada, J., Ruiz, I., & Garrido, J. (2008). Two pieces of information in a single trait? The yellow breast of the great tit (*Parus major*) reflects both pigment acquisition and body condition. *Behaviour*, 145, 1195–1210, doi:10.1163/156853908785387638.
- Shepard, R. N. (1987, Sept 9). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317–1323, doi:10.1126/science.3629243.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24, 629–640, doi:10.1017/S0140525X01000061.
- Vorobyev, M., & Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society B: Biological Sciences*, 265(1394), 351–358, doi:10.1098/rspb.1998.0302.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293–1313, doi:10.3758/Bf03194544.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63(8), 1314–1329, doi:10.3758/Bf03194545.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science* (2nd ed.). New York: Wiley.