The Auk
Ornithological Advances
AmericanOrnithology.org

Volume 133, 2016, pp. 325–337
DOI: 10.1642/AUK-16-19.1

RESEARCH ARTICLE

A quantitative analysis of objective feather color assessment: Measurements in the laboratory do not reflect true plumage color

Iker Vaquero-Alba,[1,2]* Andrew McGowan,[1] Daniel Pincheira-Donoso,[2] Matthew R. Evans,[1,a] and Sasha R. X. Dall[1]

[1] Centre for Ecology and Conservation (CEC), College of Life and Environmental Sciences, University of Exeter, Cornwall Campus, Penryn, Cornwall, UK
[2] Laboratory of Evolutionary Ecology of Adaptations, School of Life Sciences, College of Science, University of Lincoln, Brayford Campus, Lincoln, UK
[a] Current address: School of Biological and Chemical Sciences, Queen Mary, University of London, London, UK
* Corresponding author: karraspito@yahoo.es

Submitted January 21, 2016; Accepted January 25, 2016; Published April 6, 2016

ABSTRACT

An important driver of the evolution of animal coloration is sexual selection operating on traits that are used to transmit information to rivals and potential mates, which has a major impact on fitness. Reflectance spectrometry has become a standard color-measuring tool, especially after the discovery of tetrachromacy in birds and their ability to detect UV light. Birds' plumage patterns may be invisible to humans, and therefore the establishment of reliable and quantitatively objective ways of assessing coloration not dependent on human vision is a technical need of primary importance. Plumage coloration measurements can be taken directly on live birds in the field, or in the laboratory (e.g., on collected feathers). However, which of these 2 approaches offers a more reliable, repeatable sampling method remains an unsolved question. Using a spectrophotometer, we measured melanin-based coloration in the plumage of Barn Swallows (*Hirundo rustica*). We assessed the repeatability of measures obtained with both traditional sampling methods to quantitatively determine their reliability. We used an ANOVA-based method for calculating the repeatability of measurements from 2 years separately, and a GLMM-based method to calculate overall adjusted repeatabilities for both years. The results of our study indicate a great disparity between color measurements obtained using both sampling methods and a low comparability across them. Assuming that measurements taken in the field reflect the real or "true" color of plumage, we may conclude that there is a lack of reliability of the laboratory method to reflect this true color in melanin-based plumages. Likewise, we recommend the use of the GLMM-based statistical method for repeatability calculations, as it allows the inclusion of random factors and the calculation of more realistic, adjusted repeatabilities. It also reduces the number of necessary tests, thereby increasing power, and it allows easy calculation of 95% CIs, a measure of the reliability and precision of effect-size calculations.

Keywords: adjusted repeatability, bird plumage, colorful displays, sexual selection, spectrophotometry, tetrachromacy, ultraviolet

Análisis cuantitativo de la evaluación objetiva del color de las plumas: las mediciones en el laboratorio no reflejan el verdadero color del plumaje

RESUMEN

La selección sexual que opera en rasgos usados para transmitir información a rivales y potenciales parejas guía la evolución de la coloración en animales de una forma determinante. Esto tiene un impacto importante en la adecuación biológica. La espectrometría de reflectancia se ha convertido en una herramienta muy común para medir el color, especialmente tras el descubrimiento de tetracromía en las aves y su habilidad para detectar la luz ultravioleta. Los patrones del plumaje de las aves pueden ser invisibles para los humanos y, por eso, el establecimiento de formas fiables y cuantitativamente objetivas de evaluar la coloración no dependientes de la visión humana representa una necesidad técnica de vital importancia. Las mediciones de la coloración del plumaje pueden ser efectuadas directamente en aves vivas en el campo o en el laboratorio (p. ej. en plumas colectadas). Sin embargo, cuál de estos dos enfoques ofrece un método de muestreo más fiable y repetible sigue siendo una pregunta sin resolver. Usando un espectrofotómetro, medimos la coloración basada en melanina en el plumaje de la golondrina común (*Hirundo rustica*). Evaluamos la repetibilidad de las mediciones obtenidas con ambos métodos tradicionales de muestreo para determinar de modo cuantitativo su fiabilidad. Usamos el método estadístico basado en el ANOVA para calcular la repetibilidad de las mediciones de dos años por separado, y el método basado en modelos lineales mixtos generalizados (GLMM) para calcular las repetibilidades ajustadas totales para ambos años. Los resultados de nuestro

estudio indican una gran disparidad entre las mediciones de color obtenidas usando ambos métodos de muestreo y una baja comparabilidad entre ellas. Asumiendo que las mediciones efectuadas en el campo reflejan el color real o «verdadero» del plumaje, podemos concluir que hay una falta de fiabilidad del método de laboratorio para reflejar el color real de los plumajes basados en melanina. Del mismo modo, recomendamos el uso del método estadístico basado en GLMM para los cálculos de repetibilidad, ya que permite la inclusión de factores aleatorios y el cálculo de repetibilidades ajustadas, más realistas. También reduce el número de pruebas necesarias, por consiguiente aumentando la potencia estadística, y permite calcular fácilmente los intervalos de confianza (95% CI), una medida de la fiabilidad y precisión de los cálculos del tamaño del efecto.

Palabras clave: despliegues coloridos, espectrofotometría, plumaje de las aves, repetibilidad ajustada, selección sexual, tetracromía, ultravioleta

INTRODUCTION

Mate-choice theory predicts that elaborately ornamented males should provide female birds with direct (if ornamental traits reflect individual condition, useful individual attributes, or somatic quality independent of condition) and/or indirect fitness benefits ("good genes" or attractiveness for offspring, as conspicuous and costly male traits indicate highly heritable viability; Pomiankowski 1987, Andersson 1994, Garamszegi et al. 2006). Therefore, birds with more elaborate, colorful displays are expected to enjoy a selective advantage, given their higher mating chances (Andersson 1994, Hill 2006).

Color vision involves the capacity to discriminate among different wavelengths of light, independent of their intensity (Kelber et al. 2003, Cuthill 2006). Although the coloration traits that are expressed in animals have proven to be essential components to our understanding of the nature of selection, and sexual selection in particular, only relatively recently have scientists come to appreciate the importance of a systematic understanding of both the function and evolution of coloration, as well as the mechanisms that underpin it (Hill and McGraw 2006). Birds in particular, due to their colorful displays and the role of their color signals in fitness differentials, have traditionally been employed as prime model systems to understand the causes and implications of color evolution. However, the mechanisms of color vision and spectral information processing needed to understand how birds perceive colors remain areas with more questions than answers.

Two traditional ways of assessing bird plumage coloration with spectrophotometers have been reported in the literature. Measurements may be taken directly on the bird, applying the probe of the spectrophotometer to plumage patches as they occur in situ (Senar et al. 2002, Bize et al. 2006, Herrera et al. 2008, Catoni et al. 2009, Doucet and Hill 2009, Del Cerro et al. 2010). Alternatively, measurements may be taken in the lab, using feather samples collected in the field, applying the probe to "plumage patches" created by mounting these feathers on a flat surface in a way that mimics the original plumage structure (Cuthill et al. 1999, Keyser and Hill 2000, Perrier et al. 2002, McGraw et al. 2004, 2005, Safran and McGraw 2004, Komdeur et al. 2005, Safran 2007, Vaquero-Alba 2011).

Repeatability is a useful statistical tool to assess the accuracy of phenotypic measurements, as it reflects the quality of the data (Garamszegi et al. 2006). The most common measure of repeatability, or, more precisely, the coefficient of intraclass correlation (r_i), can be formally defined as the proportion of the total variance explained by differences among groups (or among measurements of the same subject); in other words, the proportion of the variance not due to measurement error or phenotypic flexibility (Lessells and Boag 1987, Nakagawa and Schielzeth 2010):

$$r_i = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2),$$

where σ_α^2 is the between-group variance and σ_ε^2 is the within-group variance, and the sum of both comprises the total phenotypic variance (Sokal and Rohlf 1995).

Despite the popularity of the use of spectrophotometers for color assessment and the growing number of studies on bird coloration, few studies have rigorously assessed the consistency of both methods for measuring the coloration of plumage patches and the repeatability of results obtained when using either one or the other. For example, Figuerola et al. (1999) analyzed the reliability of the measurements obtained by using a colorimeter to quantify the plumage coloration of 3 different passerine species. Senar et al. (2002) assessed the repeatability of the measurements taken on live birds in the field, and Perrier et al. (2002), Safran and McGraw (2004), Komdeur et al. (2005), and Safran (2007) did the same for measurements taken in the lab. Meadows et al. (2011), using Anna's Hummingbirds (*Calypte anna*) as a model species, presented a technique for measuring iridescent coloration in animals that maximizes repeatability. For most studies, however, researchers simply adopt 1 of the 2 methods to measure plumage coloration and assume that the chosen method is suitable for their purposes, but they do not really assess the reliability of the method nor try to quantify the possible measurement error that they may be incurring. To the best of our knowledge, a study by Quesada and Senar

(2006) is the only one that has actually compared both measurement methods. The authors did so using carotenoid-based plumage coloration in Eurasian Great Tits (*Parus major*). Overall, they found high values of repeatability for both methods separately and moderate-to-high values for the comparison between them.

Measuring feather coloration directly on birds in the field has practical limitations. For example, the equipment may be uncomfortable to carry into the field and fragile (Berggren and Merilä 2004), and the manipulation times of the birds may be higher. However, as patch coloration is not homogeneous, this method may better reflect the "true" variation of coloration within a patch. Collecting feathers captures just a part of the total variation, and, obviously, color in collected feathers is more homogeneous than in the whole patch. In their study, Quesada and Senar (2006) assumed that the plumage patches on the live bird were the "true" color values, and their work aimed to check whether a few collected feathers could accurately reflect that true color. It is obvious that color measurement is not totally unbiased and accurate, which is why it is important to make sure that measurements taken using both methods separately are reasonably repeatable (indicative of low measurement error). Likewise, high repeatability across methods may indicate that collected feathers accurately reflect the "true" color of the original plumage patches to a reasonable extent.

Until recently, the most common ways used to estimate the repeatabilities of data with Gaussian errors were the correlation-based method (Sokal and Rohlf 1995) and the ANOVA-based method, frequently used by behavioral and evolutionary ecologists (Donner 1986, Lessells and Boag 1987). However, Nakagawa and Schielzeth (2010) developed an innovative R-based function for calculating GLMM-based repeatability estimates, which allows for confounding variables to be factored out and calculates the confidence intervals (CIs) for each repeatability calculation, inferred from distributions of repeatabilities obtained by parametric bootstrapping. We think that this new way of calculating repeatabilities is extremely promising and may become the most widespread method used by biologists worldwide, for 2 main reasons: First, it allows for the calculation of repeatabilities from data with Gaussian and also non-Gaussian error distributions, thereby greatly widening the range of the types of data for which repeatabilities can be estimated. Second, it makes it possible to include random factors in repeatability calculations. As a result of this, adjusted repeatability values can be estimated, which are more realistic as they account for more potential sources of variance. Also, the number of necessary tests for calculating repeatability is smaller, therefore decreasing the probability of type I errors.

The aims of our study were twofold. First, we compared 2 different methods of measuring melanin-based plumage ornamentation. The methods consisted of measuring feather coloration either directly on the bird in the field or on feather samples in the lab. Second, we compared 2 statistical methods to assess repeatability itself; one method was ANOVA-based, very popular among behavioral and evolutionary ecologists and widely used, and the other was GLMM-based, newly developed and highly promising. Our first aim was to determine whether collecting feathers from birds and measuring their coloration in the lab could accurately reflect, to a reasonable extent, the true color of the plumage patch from which the feathers were collected (i.e. the values that we would get by measuring coloration directly on the bird in the field). Our second aim was to test the suitability and accuracy of the relatively new, GLMM-based statistical method, analyze its advantages, and check whether it has the potential to become the new standard method for measuring repeatability, at least in the field of behavioral and evolutionary ecology. We hypothesized that measuring coloration on feather samples in the lab would provide at least moderately representative measures of color values compared with the values obtained by measuring coloration directly on bird plumage patches, in accordance with the results obtained by Quesada and Senar (2006) for carotenoid coloration. Also, we predicted that we would achieve more realistic, as well as more robust, precise, and informative, results with the GLMM-based statistical approach, making it a very likely candidate to become widespread in the future, as it allows for the analysis of a wider range of data types and the introduction of more sources of variation into each analysis, thereby yielding more realistic repeatability estimates. We used the European subspecies of the Barn Swallow (*Hirundo rustica rustica*) as a model species. To the best of our knowledge, this is the first time that GLMM-based repeatability estimates have been used to assess the reliability of melanin-based plumage coloration measurements and to compare the reliability of measurements taken in the field vs. in the lab.

METHODS

Fieldwork and Data Collection

Plumage color measurements in the field took place between July 7 and 12 in 2009, and between June 2 and 24 in 2010. As the data collection period was very short (between 6 and 23 days), we avoided any possible color degradation effects. Measurements were carried out in multiple sites, mostly farmlands, around the Falmouth area in Cornwall, UK (Appendix Table 5). We caught 59 adult European Barn Swallows (21 in 2009 and 38 in 2010) using mist nets, banded them, took morphometric measure-

ments, quantified their plumage reflectance spectra in the field, and collected feather samples for subsequent assessment in the lab. After collection, feather samples were placed into opaque paper envelopes, and these envelopes were sealed and stored in the dark (to minimize color degradation) until the end of the field season. Color quantification of feather samples in the lab took place between November 13 and 19 in 2009, and between September 20 and 28 in 2010.

For color assessment, we used a USB2000 spectrophotometer (Ocean Optics, Dunedin, Florida, USA) and a xenon flash lamp (Ocean Optics). Before using the spectrophotometer, we calibrated it by setting the white and black references (i.e. we "told" the machine which color we wanted it to consider as the 100% reflectance (white) standard and the 0% reflectance (dark) standard), so that the rest of the color measurements were determined in relation to these maximum and minimum possible reflectance values. We used a WS-1 SS Diffuse Reflectance Standard, which is a diffuse white plastic that is >98% reflective from 250 to 1500 nm, as the white reference (100% reflectance), and a piece of black velvet as the dark standard (0% reflectance) to correct for the noise when no light reaches the sensor. At the far end of the reflection probe and light source, we put a nonreflective black sleeve. As we were interested in measuring the light that was diffusely reflected by the plumage, the sleeve was cut in a 45° angle to minimize the mismeasurement derived from the specular reflection of white light reaching the sensor (Andersson and Prager 2006). Integration time was set to 100 msec, the number of spectra averaged was 1, and electric dark correction was enabled.

Using the spectra acquisition software package OOIBase (Ocean Optics), we measured the reflectance of 4 body regions, namely the throat, breast, belly, and vent of each bird. As noted above, we measured plumage coloration on birds in the field and then collected feathers for subsequent measurement in the lab. There is quite a strong consensus regarding the delimitation and naming of the different regions or patches of birds' plumages (see, for example, Ali 1941, Forshaw 1973, Andersson and Prager 2006, and Moore et al. 2012). We used figure 2.10 from Andersson and Prager (2006) as a guide, but the limits of each region are quite consistent throughout the literature, with only some slight variation in names (some sources use "abdomen" instead of "belly," and the patch that we call the "vent" is often termed the "undertail coverts" or "crissum"). Using these standardized and perfectly delimited regions as our patches, in 2010 we placed the probe in a location as central to the patch as possible, and collected the feathers from the same point. In 2009, however, this was done differently, with measurements in the field taken in 3 separate locations within the patch, each of them approximately as far from the center of the patch as from the outer limit and as far as possible from each other, while feathers for lab measurements were collected from the center of the patch.

The number of feathers collected was always a minimum of 5, although normally ~10 whenever possible within methodological constraints. For plumage color quantification in the lab, we mounted the feathers one on top of another in an attempt to simulate the original pattern found on live birds, consistently following the method detailed in Quesada and Senar (2006; see figure 2 for a visual depiction). We mounted the feathers on a piece of black velvet to avoid background noise. In both field and lab procedures, the 2nd and 3rd measurements were made after removing the reflection probe and light source and placing them again on the color patch. I. Vaquero-Alba took all the measurements.

Due to the way in which data were collected, the 3 plumage coloration measurements taken in the field in 2009 covered a wider area of each plumage patch than the measurements made on feather samples, which were restricted to the area covered by the sample of feathers plucked from each patch of each individual. In 2010, however, the 3 field measurements were taken in approximately the same plumage area for each patch, and the feathers that were collected for lab measurements were plucked from approximately the same area from which field measurements were taken.

We used the spectral data that we obtained from the spectrophotometer to calculate brightness, chroma, and hue, parameters generally used to quantify color, as well as UV chroma, a measure of spectral purity. For the calculation of all color variables, using the equations in Endler and Mielke (2005) and the mathematical software Matlab (The MathWorks, Natick, MA, USA), we obtained the spectral sensitivity functions of the cones corrected for the cut points of oil droplets, calculated the quantal catch for each photoreceptor, and converted those quantal catches into dimensional color space coordinates in a tetrahedral color space (Figure 1). Cone sensitivities and oil droplet cut points were taken from Bowmaker et al. (1997), Vorobyev et al. (1998), Govardovskii et al. (2000), Hart (2001), and Hart and Vorobyev (2005).

Chroma is defined as the strength of the color signal or the degree of difference in stimulation among the cones, and in the tethraedral color space it is proportional to the Euclidean distance from the origin (i.e. the distance from the bird gray (achromatic) point to each point specified by 3 space coordinates). Perception of hue depends upon which cones are stimulated, and in the tetrahedral color space this is defined by the angle that a point makes with the origin. In birds, hue is defined by 2 angles, analogous to latitude and longitude in geography (Endler and Mielke 2005). We calculated brightness as the summed mean reflectance across the entire spectral range ($R_{300–700}$;

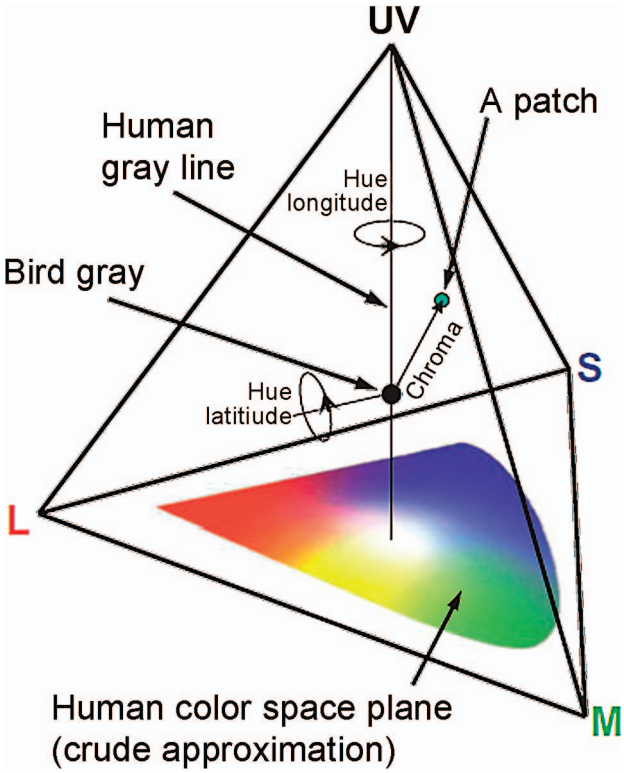

FIGURE 1. The avian tetrahedral color space (from Endler and Mielke 2005, figure 3). Permission granted by John Wiley & Sons, Ltd. © 2005 The Linnean Society of London, Biological Journal of the Linnean Society, 2005, 86, 405–431.

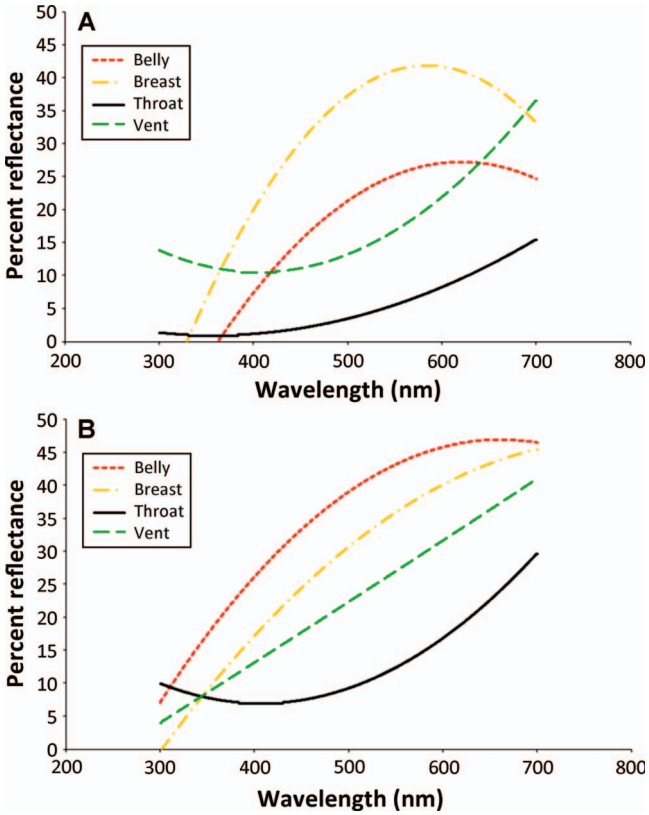

FIGURE 2. Reflectance spectra (regression lines) for belly, breast, throat, and vent patches of (**A**) male and (**B**) female Barn Swallows in Cornwall, UK, 2010.

Montgomerie 2006, Galván and Møller 2009). UV chroma was calculated as the proportion of reflectance in the UV part of the spectrum ($R_{300–400}$) in relation to the total reflectance spectrum ($R_{300–700}$; Siefferman and Hill 2005).

All of the avian families that have been investigated show plumages that reflect significant amounts of UV light (see Eaton and Lanyon [2003] for a review). In the particular case of Barn Swallows, however, Safran and McGraw (2004) reported a lack of ultraviolet reflectance for the ventral feathers of the North American subspecies (*H. r. erythrogaster*; see figure 1 in Safran and McGraw 2004). In contrast, in our population, from a different subspecies (*H. r. rustica*), although the ventral plumage showed a noisy reflectance pattern in the UV part of the spectrum, there seemed to be some reflectance in that range, with even some modest reflectance peaks for certain plumage patches (Figure 2). Therefore, we decided to include UV chroma as an additional variable in our analyses.

After the extraction of the spectral data and the calculation of the color variables, we ended up with 3 measurements per individual bird per variable per patch per method (field vs. lab). The individual repeatability of each method separately, or "within-method" repeatability, was calculated for these 3 measurements. For the assessment of the "between-method" repeatability, or comparability across methods, the 3 measurements were averaged, resulting in 2 measurements (1 for the field and 1 for the lab) per individual bird per variable per patch. The repeatability across methods was calculated for these 2 values.

We also calculated the correlation between field and lab measurements using the 2 averaged values. High correlation values would indicate that the 2 procedures measured color in a consistent way, regardless of whether absolute values were the same or not for both methods.

As we conducted multiple statistical tests on data subsets that were not likely to be biologically independent of each other (i.e. different components of the spectra, same metrics in different years, or same metrics in the lab and in the field), there was an increased probability of type I error rates. To control for this increased probability, we corrected our *P*-values for multiple tests based on the sequentially rejective Bonferroni procedure of Holm (1979) using the 'p.adjust' function in the STATS package in R (R Development Core Team 2010). All statistical analyses were carried out using R (Crawley 2007, R Development Core Team 2010).

ANOVA-based Method

We calculated repeatability for color variables in the 4 patches for the different procedures according to Lessells and Boag (1987), Senar (1999), and Quesada and Senar (2006). The individual repeatability of each method separately, or "within-method" repeatability, was computed from the mean squares of an ANOVA on 3 repeated measures per individual. For the "between-method" repeatability (i.e. the repeatability of measurements across procedures), the ANOVA was carried out on 2 repeated measures per individual, 1 from the field and 1 from the lab. We repeated this process for the 2009 and 2010 data separately.

GLMM-based Method

We used a modified version of the R function R.Anson, which is itself a modification of the rpt.remlLMM function (Nakagawa and Schielzeth 2010). We fitted 2 random-effect terms (individual identity and year) in our linear mixed-effects models, and calculated the adjusted repeatability estimate as:

$$r_i = \sigma^2_\alpha/(\sigma^2_\alpha + \sigma^2_\varepsilon + \sigma^2_r),$$

where σ^2_r is the year variance (and σ^2_α is the between-group variance and σ^2_ε is the within-group variance, as above). As for the ANOVA-based method, the repeatability of each method separately was computed for 3 repeated measures per individual, and the repeatability of measurements across procedures was calculated for 2 repeated measures per individual. Because we included year as a random effect in our models, we did not need to repeat the process for 2009 and 2010 data separately, as we did for the ANOVA method.

Although we also included Bonferroni-corrected significance levels for the GLMM-based repeatability calculations, we did this purely for the purpose of comparison with the ANOVA-method calculations. However, effect size statistics (r_i, in this case), which provide us with the magnitude of the observed effect, together with 95% CIs, which we obtained with the R.Anson function and which constitute a measure of the precision of said magnitude, are an optimal, robust, and highly informative way of presenting biological data, regardless of statistical significance, as noted in a paper (Nakagawa and Cuthill 2007) coauthored by the original author of R.Anson himself.

RESULTS

ANOVA Analyses

In 2009, measuring plumage coloration in the lab proved to be a repeatable method. Brightness, UV chroma, chroma, and hue latitude and longitude were highly repeatable for almost all of the plumage patches, returning r_i values >0.70, with the exception of hue latitude of the breast ($r_i = 0.647$, $F_{21,44} = 6.510$, $P < 0.001$), hue latitude of the throat ($r_i = 0.418$, $F_{21,44} = 3.157$, $P < 0.001$), and hue longitude of the throat ($r_i = 0.625$, $F_{21,44} = 6.012$, $P < 0.001$; Table 1).

The method of measuring plumage coloration in the field (at 3 different points, covering a wider area of each patch) was also quite consistent, but with overall lower values of repeatability, although still reasonably high. All r_i values were at least 0.50, except those for brightness of the breast ($r_i = 0.414$, $F_{20,42} = 3.118$, $P = 0.002$), hue latitude of the breast ($r_i = 0.382$, $F_{20,42} = 2.856$, $P = 0.002$), and hue latitude of the vent ($r_i = 0.394$, $F_{21,44} = 2.955$, $P = 0.001$; Table 1).

The repeatability values across the field and laboratory procedures were very low for all of the plumage patches measured ($r_i < 0.35$ and $P > 0.05$ in all cases), suggesting a lack of consistency across the 2 assessment methods for melanin-based plumage coloration (Table 1).

In 2010, repeatability measurements in the field (taken at approximately the same point within each patch) yielded considerably higher results than those from 2009, with all r_i values above 0.60, except those of hue latitude of the throat ($r_i = 0.515$, $F_{37,75} = 4.186$, $P < 0.001$). Most of the remaining values ranged from 0.74 to 0.91, except for brightness of the breast ($r_i = 0.611$, $F_{37,76} = 5.710$, $P < 0.001$), hue latitude of the belly ($r_i = 0.630$, $F_{37,76} = 6.100$, $P < 0.001$), hue latitude of the vent ($r_i = 0.629$, $F_{37,76} = 6.088$, $P < 0.001$), and hue longitude of the vent ($r_i = 0.679$, $F_{37,76} = 7.356$, $P < 0.001$; Table 2).

In the lab, all repeatability values from 2010 were higher than 0.71, except that of hue latitude of the throat ($r_i = 0.650$, $F_{37,76} = 6.569$, $P < 0.001$). Repeatability was higher overall than when taking measurements on live birds, except for UV chroma of the belly ($r_i = 0.857$, $F_{37,76} = 18.986$, $P < 0.001$), breast ($r_i = 0.788$, $F_{37,76} = 12.117$, $P < 0.001$), and vent ($r_i = 0.819$, $F_{37,76} = 14.571$, $P < 0.001$), and hue latitude of the breast ($r_i = 0.722$, $F_{37,76} = 8.808$, $P < 0.001$), where it was just slightly lower. Repeatability values in the lab were similar to the results obtained in 2009, but visibly higher for measurements taken in the field (except for throat patch; Table 2).

Repeatabilities across field and lab methods in 2010 were quite heterogeneous: high for hue longitude of the belly ($r_i = 0.794$, $F_{37,38} = 8.732$, $P < 0.001$) and breast ($r_i = 0.657$, $F_{37,38} = 4.818$, $P < 0.001$); moderate for vent hue latitude ($r_i = 0.463$, $F_{37,38} = 2.723$, $P = 0.005$) and longitude ($r_i = 0.561$, $F_{37,38} = 3.553$, $P < 0.001$), belly hue latitude ($r_i = 0.431$, $F_{37,38} = 2.515$, $P = 0.011$), and breast brightness ($r_i = 0.482$, $F_{37,38} = 2.861$, $P = 0.003$); and low for breast chroma ($r_i = 0.326$, $F_{37,38} = 1.966$, $P = 0.062$), throat UV chroma ($r_i = 0.321$, $F_{37,38} = 1.944$, $P = 0.112$), and vent brightness ($r_i = 0.349$, $F_{37,38} = 2.070$, $P = 0.042$). For the rest of the coloration measurements, repeatabilities were very low ($r_i < 0.30$ and $P > 0.05$ in all cases; Table 2).

TABLE 1. ANOVA-derived repeatabilities for 2009 plumage coloration measurements taken from live Barn Swallows in the field (in Cornwall, UK), from feather samples in the lab, and across both procedures.

	Belly		Breast		Throat		Vent	
	F	r_i	*F*	r_i	*F*	r_i	*F*	r_i
Repeatability, field								
Brightness	6.045	0.627***	3.118	0.414**	12.273	0.790***	6.187	0.634***
UV chroma	10.203	0.754***	14.171	0.814***	13.861	0.811***	5.320	0.590***
Chroma	9.016	0.728***	11.238	0.773***	10.714	0.764***	4.561	0.543***
Hue latitude	4.092	0.508***	2.856	0.382**	4.415	0.532***	2.955	0.394**
Hue longitude	13.025	0.800***	9.142	0.731***	24.348	0.886***	4.677	0.550***
Repeatability, lab								
Brightness	13.188	0.802***	14.777	0.821***	8.209	0.706***	20.212	0.865***
UV chroma	46.493	0.938***	23.015	0.880***	34.895	0.919***	25.561	0.892***
Chroma	42.489	0.932***	26.975	0.896***	62.481	0.954***	29.052	0.903***
Hue latitude	11.986	0.785***	6.510	0.647***	3.157	0.418***	23.024	0.880***
Hue longitude	25.986	0.893***	9.283	0.734***	6.012	0.625***	27.291	0.898***
Comparison, field–lab								
Brightness	1.464	0.188	2.070	0.349	0.865	–0.072	1.804	0.287
UV chroma	1.315	0.136	1.359	0.152	0.750	–0.143	0.901	–0.052
Chroma	0.327	–0.507	1.372	0.157	1.014	0.007	0.138	–0.758
Hue latitude	0.908	–0.048	0.864	–0.073	1.252	0.112	0.706	–0.172
Hue longitude	1.736	0.269	1.751	0.273	0.671	–0.197	1.106	0.050

*** $P < 0.001$; ** $P < 0.01$.

GLMM Analyses

When considering adjusted repeatability for both sampling years using the GLMM method, r_i values in the field were moderate overall, with all of them >0.50 except those of chroma of the throat ($r_i = 0.457$, 95% CI = 0.152–0.849), hue latitude of the throat ($r_i = 0.444$, 95% CI = 0.232–0.627), and hue latitude of the vent ($r_i = 0.446$, 95% CI = 0.248–0.612). The rest of the r_i values in general ranged between 0.512 and 0.669, with moderately broad 95% CIs. Only 3 values were above 0.70 and had quite narrow 95% CIs: hue longitude of the belly ($r_i = 0.858$, 95% CI = 0.768–0.905), hue longitude of the breast ($r_i = 0.840$, 95% CI = 0.744–0.890), and hue longitude of the throat ($r_i = 0.776$, 95% CI = 0.550–0.885; Table 3).

TABLE 2. ANOVA-derived repeatabilities for 2010 plumage coloration measurements taken from live Barn Swallows in the field (in Cornwall, UK), from feather samples in the lab, and across both procedures.

	Belly		Breast		Throat		Vent	
	F	r_i	*F*	r_i	*F*	r_i	*F*	r_i
Repeatability, field								
Brightness	16.422	0.837***	5.710	0.611***	17.398	0.846***	14.534	0.819***
UV chroma	25.333	0.890***	15.853	0.832***	12.357	0.791***	16.829	0.841***
Chroma	24.037	0.885***	14.749	0.821***	22.212	0.876***	12.377	0.792***
Hue latitude	6.100	0.630***	9.741	0.744***	4.186	0.515***	6.088	0.629***
Hue longitude	23.669	0.883***	31.363	0.910***	9.891	0.748***	7.356	0.679***
Repeatability, lab								
Brightness	55.197	0.947***	31.387	0.910***	30.900	0.909***	44.036	0.934***
UV chroma	18.986	0.857***	12.117	0.788***	17.544	0.847***	14.571	0.819***
Chroma	25.375	0.890***	25.936	0.893***	27.679	0.899***	39.854	0.928***
Hue latitude	8.391	0.711***	8.808	0.722***	6.569	0.650***	9.142	0.731***
Hue longitude	37.357	0.924***	31.683	0.911***	11.664	0.781***	37.517	0.924***
Comparison, field–lab								
Brightness	1.304	0.132	2.861	0.482**	0.764	–0.134	2.070	0.349*
UV chroma	0.557	–0.284	1.059	0.029	1.944	0.321	1.587	0.227
Chroma	0.906	–0.049	1.966	0.326 §	1.755	0.274	1.579	0.224
Hue latitude	2.515	0.431*	1.248	0.110	0.784	–0.121	2.723	0.463**
Hue longitude	8.732	0.794***	4.818	0.657***	0.973	–0.014	3.553	0.561***

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; § $P < 0.1$.

TABLE 3. GLMM-derived repeatabilities for years 2009 and 2010 combined for plumage coloration measurements taken from live Barn Swallows in the field (in Cornwall, UK), from feather samples in the lab, and across both procedures.

	Belly		Breast		Throat		Vent	
	r_i	95% CL	r_i	95% CL	r_i	95% CL	r_i	95% CL
Repeatability, field								
Brightness	0.591***	[0.289, 0.813]	0.512***	[0.319, 0.664]	0.664***	[0.382, 0.834]	0.669***	[0.440, 0.806]
UV chroma	0.518***	[0.194, 0.891]	0.647***	[0.319, 0.869]	0.593***	[0.263, 0.847]	0.640***	[0.353, 0.817]
Chroma	0.528***	[0.194, 0.878]	0.624***	[0.310, 0.851]	0.457***	[0.152, 0.849]	0.551***	[0.265, 0.773]
Hue latitude	0.548***	[0.367, 0.686]	0.632***	[0.471, 0.740]	0.444***	[0.232, 0.627]	0.446***	[0.248, 0.612]
Hue longitude	0.858***	[0.768, 0.905]	0.840***	[0.744, 0.890]	0.776***	[0.550, 0.885]	0.602***	[0.421, 0.728]
Repeatability, lab								
Brightness	0.677***	[0.341, 0.907]	0.853***	[0.710, 0.913]	0.829***	[0.729, 0.882]	0.853***	[0.665, 0.929]
UV chroma	0.882***	[0.797, 0.921]	0.818***	[0.694, 0.883]	0.668***	[0.331, 0.895]	0.744***	[0.458, 0.885]
Chroma	0.897***	[0.795, 0.939]	0.898***	[0.830, 0.931]	0.926***	[0.872, 0.950]	0.911***	[0.849, 0.940]
Hue latitude	0.748***	[0.617, 0.824]	0.641***	[0.440, 0.766]	0.544***	[0.375, 0.662]	0.414***	[0.142, 0.768]
Hue longitude	0.896***	[0.807, 0.934]	0.631***	[0.282, 0.892]	0.472***	[0.189, 0.769]	0.602***	[0.248, 0.911]
Comparison, field–lab								
Brightness	0.114	[0.000, 0.347]	0.441***	[0.216, 0.628]	0.000	[0.000, 0.240]	0.311*	[0.063, 0.511]
UV chroma	0.000	[0.000, 0.238]	0.085	[0.000, 0.320]	0.057	[0.000, 0.294]	0.140	[0.000, 0.364]
Chroma	0.039	[0.000, 0.273]	0.249*	[0.023, 0.470]	0.150	[0.000, 0.374]	0.213	[0.000, 0.430]
Hue latitude	0.180	[0.000, 0.408]	0.083	[0.000, 0.314]	0.000	[0.000, 0.233]	0.104	[0.000, 0.328]
Hue longitude	0.656***	[0.469, 0.778]	0.501***	[0.270, 0.697]	0.000	[0.000, 0.238]	0.316**	[0.109, 0.553]

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

Adjusted repeatability values obtained in the lab were considerably higher than those achieved in the field. All lab r_i values were >0.60, except for hue latitude of the throat ($r_i = 0.544$, 95% CI = 0.375–0.662), hue longitude of the throat ($r_i = 0.472$, 95% CI = 0.189–0.769), and hue latitude of the vent ($r_i = 0.414$, 95% CI = 0.142–0.768). Of the remaining r_i values, most ranged from 0.602 to 0.898, with chroma of the throat ($r_i = 0.926$, 95% CI = 0.872–0.950) and chroma of the vent ($r_i = 0.911$, 95% CI = 0.849–0.940) even exceeding 0.90 and with extremely narrow 95% CIs (Table 3).

The comparison between coloration measurements obtained in the field and in the lab yielded highly heterogeneous results: moderate to high repeatability values for hue longitude of the belly ($r_i = 0.656$, 95% CI = 0.469–0.778) and the breast ($r_i = 0.501$, 95% CI = 0.270–0.697); low to moderate values for brightness of the breast ($r_i = 0.441$, 95% CI = 0.216–0.628); and low values for brightness of the vent ($r_i = 0.311$, 95% CI = 0.063–0.511) and hue longitude of the vent ($r_i = 0.316$, 95% CI = 0.109–0.553). The rest of the r_i values were very low (<0.25; Table 3).

Correlation Analyses

Correlations between coloration values obtained in the field and in the lab followed very similar patterns as repeatabilities across both methods. In 2009, correlation coefficients were low to very low and nonsignificant in all cases.

In 2010, the correlation between field and lab measurements was high for hue latitude ($\rho_{x,y} = 0.678$, $t_{36} = 5.539$, $P < 0.001$) and hue longitude of the belly ($\rho_{x,y} = 0.840$, $t_{36} = 9.273$, $P < 0.001$), and for hue longitude of the breast ($\rho_{x,y} = 0.766$, $t_{36} = 7.148$, $P < 0.001$); moderate for brightness of the breast ($\rho_{x,y} = 0.473$, $t_{36} = 3.219$, $P = 0.005$), UV chroma of the throat ($\rho_{x,y} = 0.412$, $t_{36} = 2.715$, $P = 0.025$), hue latitude ($\rho_{x,y} = 0.497$, $t_{36} = 3.439$, $P = 0.003$) and hue longitude of the vent ($\rho_{x,y} = 0.593$, $t_{36} = 4.424$, $P < 0.001$); and low (<0.36) in all other cases (Table 4).

When including both years in the analyses, correlation coefficients were very similar to the ones from 2010, with only slightly lower absolute values, and significance levels were also very close, with the exception of UV chroma of the throat ($\rho_{x,y} = 0.237$, $t_{58} = 1.861$, $P = 0.161$) and hue latitude of the vent ($\rho_{x,y} = 0.218$, $t_{58} = 1.703$, $P = 0.094$), for which correlation coefficients were considerably lower than in 2010 and nonsignificant (Table 4).

DISCUSSION

Measuring Plumage Ornamentation to Gain Repeatable and Reliable Measures

Color measurements in the field and in the lab were in general moderately to highly repeatable for all of the variables and plumage patches that we examined in 2009, in 2010, and when applying the GLMM-based method for both years combined, with just some occasional exceptions. Repeatability for field measurements was visibly

TABLE 4. Correlation values between field and lab measurements of plumage coloration of Barn Swallows in Cornwall, UK, in 2009, in 2010, and for both years simultaneously.

	Belly		Breast		Throat		Vent	
	t	$\rho_{x,y}$	*t*	$\rho_{x,y}$	*t*	$\rho_{x,y}$	*t*	$\rho_{x,y}$
Correlation, field–lab (2009)								
Brightness	0.728	0.165	1.561	0.337	0.818	0.180	1.235	0.266
UV chroma	0.570	0.130	0.602	0.137	1.207	0.261	0.453	0.101
Chroma	1.076	0.240	0.633	0.144	0.444	0.099	1.467	0.312
Hue latitude	0.123	0.028	0.753	0.170	0.533	0.118	–0.870	–0.191
Hue longitude	1.128	0.251	1.907	0.401	–0.103	–0.023	0.121	0.027
Correlation, field–lab (2010)								
Brightness	0.738	0.122	3.219	0.473**	1.803	0.288	2.273	0.354*
UV chroma	–1.854	–0.295	0.173	0.029	2.715	0.412*	1.419	0.230
Chroma	–0.322	–0.054	1.999	0.316 §	2.228	0.348 §	1.432	0.232
Hue latitude	5.539	0.678***	1.947	0.309 §	0.271	0.045	3.439	0.497**
Hue longitude	9.273	0.840***	7.148	0.766***	1.243	0.203	4.424	0.593***
Correlation, field–lab (both years)								
Brightness	1.003	0.132	3.847	0.454***	1.885	0.240	2.592	0.322*
UV chroma	–1.211	–0.158	0.697	0.092	1.861	0.237	1.234	0.160
Chroma	0.429	0.057	2.481	0.312*	1.645	0.211	2.008	0.255 §
Hue latitude	3.771	0.447***	2.513	0.316*	0.374	0.049	1.703	0.218 §
Hue longitude	7.302	0.695***	7.277	0.694***	0.350	0.046	3.936	0.459**

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; § $P < 0.1$.

lower in 2009, whereas lab measurements were similarly repeatable in both years.

Comparability across both measuring methods, however, was highly heterogeneous and poor overall, with extremely low and nonsignificant r_i values in 2009. Some of the values were $\geq$0.5 and statistically significant in 2010, but, as in 2009, were also mostly low and nonsignificant.

We can partly explain the higher repeatability of 2010 field measurements and the higher frequency of large and significant repeatability values in the comparison between field and lab methods in 2010 by the different methodology that we followed in field data collection between years, as explained above. However, the fact that most of the repeatability values for field–lab comparisons in 2010 were still very low indicates that there must be additional and more influential factors to explain these findings. Our results stand in marked contrast to the positive results of Quesada and Senar (2006), who compared the repeatabilities between both coloration assessment procedures for carotenoid-based plumage. There may be several reasons for this difference. For example, due to the different characteristics of the 2 types of pigment, carotenoid-derived coloration is more variable among individuals than melanin-based coloration (Badyaev and Hill 2000), and the repeatability of a character increases with between-individual (between-group) variability (Senar 1999). In order to increase the repeatability of some measurements, a possible solution could be to increase the number of measurements, for example from 3 to 5, as has already been done by several authors (Bennett et al. 1997, Perrier et al. 2002, Doucet and Hill 2009). However, when working with live birds in the field, by doing this we would increase manipulation times and, consequently, we might increase stress levels to what we would consider an unacceptable degree according to our experience with Barn Swallows. However, this solution could be applied when assessing coloration using feather samples in the lab.

The comparability across methods was especially low for measurements taken from the throat, even when feathers were collected from the same approximate point where field measurements were made (in 2010) or when considering the adjusted repeatability for both years. A potential explanation for this is that the throat patch is smaller and much darker than other plumage patches. The feathers of the throat patch are also considerably smaller. Therefore, it is often quite difficult to obtain a reliable reflectance measurement with such a limited number of photons reaching the spectrophotometer probe. Also, it is more difficult to create a "plumage patch" in the lab with a feather arrangement similar to that of a live bird and big enough to apply a spectrophotometer probe to.

The fact that the correlation coefficients between field and lab measurements followed a pattern closely related to the repeatability values across both procedures suggests that, in broad terms, both sampling methods neither yielded close and comparable absolute values for color measurements, nor did they even measure color in a consistent way. Thus, according to our evidence, collecting feathers from birds and quantifying their coloration in the lab does not accurately or reliably reflect the true color of

the plumage patches that can be perceived in the wild and from which the feathers were collected, at least for plumages with melanin-based pigmentation. This stands in contrast to the results obtained by Quesada and Senar (2006) for carotenoid-based plumages. Thus, we do not recommend measuring coloration in the lab for consistent, reliable, and objective feather color assessment. On the contrary, we recommend carrying the spectrophotometer into the field, as we believe that the reliability of the color measurements taken on live birds clearly outweighs the related inconveniences.

Assessing Repeatability

The GLMM-based method (Nakagawa and Schielzeth 2010), applied to data from both years combined, allowed us to control for year effects by adding the year variance into the total variance calculation, so that we could obtain the adjusted repeatability for data from both years. In addition, use of the GLMM-based method meant that we could calculate 95% CIs, useful indicators of the reliability and precision of our repeatability estimates, which, together with the effect size statistic itself (r_i), provided us with a robust and highly informative way of presenting the repeatability data independent of the statistical significance level (Nakagawa and Cuthill 2007). Further, and despite the lack of a need for *P*-values, the possibility of calculating adjusted repeatabilities by including, in our case, year as a random effect considerably reduced the number of tests necessary for repeatability calculation. Similarly, any other possible random factor potentially having an effect on the repeatability of measurements can be included in analyses, thereby greatly reducing the number of statistical tests required. Thus, the *P*-values obtained with this method are less affected by problems derived from multiple testing than those obtained with the ANOVA-based method, reducing the probability of type I errors and increasing the power of this repeatability-calculation method.

Although, with some infrequent exceptions, the absolute values of repeatability were higher overall for lab than for field measurements within years when using the ANOVA approach, and also when considering data from both years simultaneously when using the GLMM-based procedure, this does not necessarily mean that color measurements taken in the lab are more unbiased and accurate. It simply indicates that collecting feathers captures just a portion of the total variation of the plumage patch from which they were collected. Color across collected feathers is, by definition, more homogeneous than that in the true patch. Even if feathers are collected from exactly the same point at which the field measurements are taken, as we tried to do in 2010, the subsequent stacking of the feathers 1 by 1 for measurement in the lab may alter their original arrangement and cause a reduction in the comparability across methods. Likewise, the number of feathers that are stacked for lab measurements, which we failed to control, is a crucial variable with a potentially dramatic effect on results. We still think, however, that the unique characteristics of melanin-based pigmentation make the plumage color of Barn Swallows especially complicated to quantify in a reliable manner and have a detrimental effect on repeatability. Regardless, for future work on this topic, we strongly suggest that special care is taken to control the following 2 aspects of paramount importance: collecting feathers from the same point at which direct measurements in the field are taken as exactly as possible, and controlling for the number of feathers stacked for lab measurements. By doing this, we predict that there can be an increase in the comparability across field and lab color quantification procedures. We believe that finding out to what degree this increases comparability is a promising and exciting question to investigate, and may give us much information about the convenience of using the lab procedure for color quantification in melanin-based plumage, given that in view of the available data so far we cannot recommend its use.

Conclusions

The results of our study suggest that, in species with melanin-based plumage coloration, the procedure of collecting feathers from a bird, stacking them on a flat surface in the lab while trying to mimic the original plumage arrangement, and quantifying their coloration with a spectrophotometer does not reliably reflect the real or true color of the bird's plumage in the field. Therefore, we recommend quantifying coloration directly in the field by carrying the spectrophotometer to the site for accurate, reliable, and realistic color assessment.

Likewise, we advocate the use of the GLMM-based statistical method for repeatability calculations, as it allows for the inclusion of random factors in models. Consequently, adjusted, highly realistic repeatabilities can be calculated. Also, the number of statistical tests necessary for repeatability calculation is reduced, increasing power. Finally, this method allows the easy computation of 95% CIs, a measure of the reliability and precision of effect size calculations. In sum, the GLMM-based method constitutes a realistic, robust, highly informative, reliable, and accurate way of calculating repeatability and presenting data, and we recommend its generalized use, instead of use of the more popular and common methods employed to date.

ACKNOWLEDGMENTS

Thanks to Mary-Anne Collis, who contributed to establishing the farm network, and Ian Blessley, who extended the network

and initially talked to some of the farmers. Our gratitude to Shinichi Nakagawa, who helped us with the modification of his R function for our purposes. Rebecca Safran was a huge inspiration for our work. Jon Blount introduced I.V.-A. to the use of the spectrophotometer, and Tom Pike allowed us to use his Matlab spectral data processing equations. Thanks to Joan Carles Senar for his invaluable help with repeatability-related issues. We are very much indebted to all of the farmers and landlords who granted us access to their properties, their lovely families, neighbors, and pets. Two reviewers provided helpful comments on a previous version of the manuscript.

Funding statement: I.V.-A. was sponsored by a PhD studentship from the Programa de Formación de Personal Investigador, Departamento de Educación, Universidades e Investigación, Gobierno Vasco, Spain. The funder did not have any input into the content of the manuscript, nor require approval of the manuscript before submission or publication.

Ethics statement: A.M. and M.R.E. had a Home Office license, which covered taking feather samples as well as other activities (Home Office Project License Number 30/2740). M.R.E. was the project license holder. All work was carried out on private residencies and farms with the express permission of the landowners in question. Contact details of the landowners can be provided by the authors upon request and after asking the landowners for permission, in order to respect their privacy. The specific locations of the study are provided in Appendix Table 5. Our field study did not involve any endangered or protected species. Birds were caught using mist nets under license (A.M., BTO A License Holder No. 4947).

Author contributions: I.V.-A. collected and analyzed the data, formulated hypotheses, developed statistical methods, and wrote the paper. A.M. supervised field work and data collection, designed field methods, and contributed substantial resources. D.P.-D. conceived the idea of the paper, supervised the theoretical aspects, and substantially edited the paper. S.R.X.D. and M.R.E. supervised fieldwork, data collection, and research in general, designed methods, and edited the paper.

LITERATURE CITED

Ali, S. (1941). The Book of Indian Birds. The Bombay Natural History Society, Bombay, India.

Andersson, M. B. (1994). Sexual Selection. Princeton University Press, Princeton, NJ, USA.

Andersson, S., and M. Prager (2006). Quantifying colors. In Bird Coloration, volume I: Mechanisms and Measurements (G. E. Hill and K. J. McGraw, Editors). Harvard University Press, Cambridge, MA, USA. pp. 41–89.

Badyaev, A. V., and G. E. Hill (2000). Evolution of sexual dichromatism: Contribution of carotenoid- versus melanin-based coloration. Biological Journal of the Linnean Society 69:153–172.

Bennett, A. T. D., I. C. Cuthill, J. C. Partridge, and K. Lunau (1997). Ultraviolet plumage colors predict mate preferences in starlings. Proceedings of the National Academy of Sciences USA 94:8618–8621.

Berggren, Å., and J. Merilä (2004). WWW design code—A new tool for colour estimation in animal studies. Frontiers in Zoology 1:2. doi:10.1186/1742-9994-1-2

Bize, P., R. Piault, B. Moureau, and P. Heeb (2006). A UV signal of offspring condition mediates context-dependent parental favouritism. Proceedings of the Royal Society of London, Series B 273:2063–2068.

Bowmaker, J. K., L. A. Heath, S. E. Wilkie, and D. M. Hunt (1997). Visual pigments and oil droplets from six classes of photoreceptor in the retinas of birds. Vision Research 37: 2183–2194.

Catoni, C., A. Peters, and H. M. Schaefer (2009). Dietary flavonoids enhance conspicuousness of a melanin-based trait in male Blackcaps but not of the female homologous trait or of sexually monochromatic traits. Journal of Evolutionary Biology 22:1649–1657.

Crawley, M. J. (2007). The R Book. Wiley Publishing, Chichester, UK.

Cuthill, I. C. (2006). Color perception. In Bird Coloration, volume I: Mechanisms and Measurements (G. E. Hill and K. J. McGraw, Editors). Harvard University Press, Cambridge, MA, USA. pp. 3–40.

Cuthill, I. C., A. T. D. Bennett, J. C. Partridge, and E. J. Maier (1999). Plumage reflectance and the objective assessment of avian sexual dichromatism. American Naturalist 153:183–200.

Del Cerro, S., S. Merino, J. Martínez-de la Puente, E. Lobato, R. Ruiz-de-Castañeda, J. Rivero-de Aguilar, J. Martínez, J. Morales, G. Tomás, and J. Moreno (2010). Carotenoid-based plumage colouration is associated with blood parasite richness and stress protein levels in Blue Tits (*Cyanistes caeruleus*). Oecologia 162:825–835.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. International Statistical Review/Revue Internationale de Statistique 54:67–82.

Doucet, S. M., and G. E. Hill (2009). Do museum specimens accurately represent wild birds? A case study of carotenoid, melanin, and structural colours in Long-tailed Manakins *Chiroxiphia linearis*. Journal of Avian Biology 40:146–156.

Eaton, M. D., and S. M. Lanyon (2003). The ubiquity of avian ultraviolet plumage reflectance. Proceedings of the Royal Society of London, Series B 270:1721–1726.

Endler, J. A., and P. W. Mielke (2005). Comparing entire color patterns as birds see them. Biological Journal of the Linnean Society 86:405–431.

Figuerola, J., J. C. Senar, and J. Pascual (1999). The use of a colorimeter in field studies of Blue Tit *Parus caeruleus* coloration. Ardea 87:269–275.

Forshaw, J. M. (1973). Parrots of the World. Lansdowne Press, Willoughby, NSW, Australia.

Galván, I., and A. P. Møller (2009). Different roles of natural and sexual selection on senescence of plumage color in the Barn Swallow. Functional Ecology 23:302–309.

Garamszegi, L. Z., G. Hegyi, D. Heylen, P. Ninni, F. de Lope, M. Eens, and A. P. Møller (2006). The design of complex sexual traits in male Barn Swallows: Associations between signal attributes. Journal of Evolutionary Biology 19:2052–2066.

Govardovskii, V. I., N. Fyhrquist, T. Reuter, D. G. Kuzmin, and K. Donner (2000). In search of the visual pigment template. Visual Neuroscience 17:509–528.

Hart, N. S. (2001). The visual ecology of avian photoreceptors. Progress in Retinal and Eye Research 20:675–703.

Hart, N. S., and M. Vorobyev (2005). Modelling oil droplet absorption spectra and spectral sensitivities of bird cone photoreceptors. Journal of Comparative Physiology A 191: 381–392.

Herrera, G., J. C. Zagal, M. Diaz, M. J. Fernández, A. Vielma, M. Cure, J. Martinez, F. Bozinovic, and A. G. Palacios (2008). Spectral sensitivities of photoreceptors and their role in color discrimination in the Green-backed Firecrown Hummingbird (*Sephanoides sephaniodes*). Journal of Comparative Physiology A 194:785–794.

Hill, G. E. (2006). Female mate choice for ornamental coloration. In Bird Coloration, volume II: Function and Evolution (G. E. Hill and K. J. McGraw, Editors). Harvard University Press, Cambridge, MA, USA. pp. 2137–2200.

Hill, G. E., and K. J. McGraw (Editors) (2006). Bird Coloration, volume I: Mechanisms and Measurements. Harvard University Press, Cambridge, MA, USA.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65–70.

Kelber, A., M. Vorobyev, and D. Osorio (2003). Animal color vision—Behavioural tests and physiological concepts. Biological Reviews 78:81–118.

Keyser, A. J., and G. E. Hill (2000). Structurally based plumage coloration is an honest signal of quality in male Blue Grosbeaks. Behavioral Ecology 11:202–209.

Komdeur, J., M. Oorebeek, T. van Overveld, and I. C. Cuthill (2005). Mutual ornamentation, age, and reproductive performance in the European Starling. Behavioral Ecology 16:805–817.

Lessells, C. M., and P. T. Boag (1987). Unrepeatable repeatabilities: A common mistake. The Auk 104:116–121.

McGraw, K. J., R. J. Safran, M. R. Evans, and K. Wakamatsu (2004). European Barn Swallows use melanin pigments to color their feathers brown. Behavioral Ecology 15:889–891.

McGraw, K. J., R. J. Safran, and K. Wakamatsu (2005). How feather color reflects its melanin content. Functional Ecology 19:816–821.

Meadows, M. G., N. I. Morehouse, R. L. Rutowski, J. M. Douglas, and K. J. McGraw (2011). Quantifying iridescent coloration in animals: A method for improving repeatability. Behavioral Ecology and Sociobiology 65:1317–1327.

Montgomerie, R. (2006). Analyzing colors. In Bird Coloration, volume I: Mechanisms and Measurements (G. E. Hill and K. J. McGraw, Editors). Harvard University Press, Cambridge, MA, USA. pp. 90–147.

Moore, B. A., P. Baumhardt, M. Doppler, J. Randolet, B. F. Blackwell, T. L. DeVault, E. R. Loew, and E. Fernández-Juricic (2012). Oblique color vision in an open-habitat bird: Spectral sensitivity, photoreceptor distribution and behavioral implications. Journal of Experimental Biology 215: 3442–3452.

Nakagawa, S., and I. C. Cuthill (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. Biological Reviews 82:591–605.

Nakagawa, S., and H. Schielzeth (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. Biological Reviews 85:935–956.

Perrier, C., F. de Lope, A. P. Møller, and P. Ninni (2002). Structural coloration and sexual selection in the Barn Swallow *Hirundo rustica*. Behavioral Ecology 13:728–736.

Pomiankowski, A. (1987). Sexual selection: The handicap principle does work – sometimes. Proceedings of the Royal Society of London, Series B 231:123–145.

Quesada, J., and J. C. Senar (2006). Comparing plumage color measurements obtained directly from live birds and from collected feathers: The case of the Great Tit *Parus major*. Journal of Avian Biology 37:609–616.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Safran, R. J. (2007). Settlement patterns of female Barn Swallows *Hirundo rustica* across different group sizes: Access to colorful males or favored nests? Behavioral Ecology and Sociobiology 61:1359–1368.

Safran, R. J., and K. J. McGraw (2004). Plumage coloration, not length or symmetry of tail-streamers, is a sexually selected trait in North American Barn Swallows. Behavioral Ecology 15: 455–461.

Senar, J. (1999). La medición de la repetibilidad y el error de medida. EtoloGuía 17:53–64.

Senar, J. C., J. Figuerola, and J. Pascual (2002). Brighter yellow Blue Tits make better parents. Proceedings of the Royal Society of London, Series B 269:257–261.

Siefferman, L., and G. E. Hill (2005). Evidence for sexual selection on structural plumage coloration in female Eastern Bluebirds (*Sialia sialis*). Evolution 59:1819–1828.

Sokal, R. R., and F. J. Rohlf (1995). Biometry: The Principles and Practice of Statistics in Biological Research. W. H. Freeman, New York, NY, USA.

Vaquero-Alba, I. (2011). Animal sexual signals: Do they maximise or optimise information content? Ph.D. dissertation, University of Exeter, Penryn, UK.

Vorobyev, M., D. Osorio, A. T. D. Bennett, N. J. Marshall, and I. C. Cuthill (1998). Tetrachromacy, oil droplets and bird plumage colours. Journal of Comparative Physiology A 183:621–633.

Appendix Table 5. Specific locations of the study sites (GPS coordinates) in Cornwall, UK, where we took plumage coloration measurements of Barn Swallows in the field and collected feather samples for subsequent coloration measurements in the lab, assessment of repeatability of measurements, and comparability between field and lab measurements.

Site	Latitude	Longitude
Boskensoe Farm	50°07′08′′N	05°06′54′′W
Treworval Farm	50°07′16′′N	05°08′02′′W
Stithians Reservoir	50°10′05′′N	05°12′09′′W
Gwen Chapel	50°10′23′′N	05°12′45′′W
Halabezack Farm	50°00′25′′N	05°13′00′′W
Upper Menherion Farm	50°11′19′′N	05°13′30′′W
Crowgey	50°12′08′′N	05°11′50′′W
Higher Trevethan	50°13′34′′N	05°11′29′′W
Dougie Grahams	50°10′32′′N	05°06′24′′W
Porloe Farm	50°10′29′′N	05°03′34′′W
Little Tregew	50°10′45′′N	05°04′31′′W
Lower Treluswell Farm	50°11′00′′N	05°06′56′′W
Dowstall Farm	50°11′33′′N	05°04′59′′W
Pellynwartha Farm	50°12′16′′N	05°08′19′′W
Restronguet Barton Farm 1	50°11′15′′N	05°03′47′′W
Restronguet Barton Farm 2	50°11′26′′N	05°04′03′′W
University of Exeter Cornwall Campus	50°10′15′′N	05°07′27′′W
Tremough Farm	50°10′06′′N	05°07′31′′W
Rosehill Farm	50°10′59′′N	05°04′43′′W
Girl Guides Camp, St Clements	50°15′27′′N	05°00′50′′W
Polquick Farm	50°16′24′′N	05°03′07′′W