J Ornithol (2014) 155:701–706
DOI 10.1007/s10336-014-1053-9

ORIGINAL ARTICLE

Visual scoring of eggshell patterns has poor repeatability

Kaat Brulez · Pankaj K. Choudhary · Golo Maurer · Steven J. Portugal · Rebecca L. Boulton · Simone L. Webber · Phillip Cassey

Received: 27 August 2013 / Revised: 12 November 2013 / Accepted: 4 February 2014 / Published online: 18 February 2014

Abstract Eggshell pattern scoring, a method to quantify the degree of surface maculation, can potentially be a quick, inexpensive and reliable method to obtain information on eggshell appearance and spot patterns. The key pigment responsible for red-brownish hues, protoporphyrin IX, is often localized as spots, either on the surface or in distinct layers within the eggshell. Heritable pigment spotting has been linked to factors such as breeding performance and eggshell strength. In this study, we investigated whether pigment scoring of eggshell patterns is repeatable within and between observers, by testing observers under standardised conditions, using the eggshells of two commonly studied passerines, Great Tits (*Parus major*) and Blue Tits (*Cyanistes caeruleus*). We found that repeatability of eggshell scores was poor, both within and between observers for both the species. We, therefore, encourage future studies to use alternative methods for quantifying spot patterns, such as digital image analysis, a technique which has already been used extensively.

Keywords Pigmentation · Protoporphyrin · Repeatability · Spot scoring · Tits

Communicated by F. Bairlein.

Electronic supplementary material The online version of this article (doi:10.1007/s10336-014-1053-9) contains supplementary material, which is available to authorized users.

K. Brulez (✉) · G. Maurer · S. J. Portugal · R. L. Boulton · S. L. Webber · P. Cassey
Centre for Ornithology, School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK
e-mail: kaat.brulez@gmail.com

P. K. Choudhary
Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083-0688, USA

G. Maurer · R. L. Boulton · P. Cassey
School of Earth and Environmental Sciences, University of Adelaide, Adelaide, SA 5005, Australia

S. J. Portugal
Structure and Motion Lab, Royal Veterinary College, University of London, Hatfield, Hertfordshire AL9 7TA, UK

S. L. Webber
NHBS Ltd., 2-3 Wills Road, Totnes, Devon TQ9 5XN, UK

Zusammenfassung

Optische Kategorisierung von Eierschalenmustern besitzt nur geringe Wiederholpräzision

Die Kategorisierung von Eierschalenmustern, eine Methode zur Quantifizierung der Intensität der Oberflächenfleckung, kann eine schnelle, kostengünstige und zuverlässige Methode darstellen, um Informationen über Eierschalenerscheinung und Fleckenverteilung zu erlangen. Das Pigment, das hauptsächlich für die rot-braunen Farbtöne verantwortlich ist, das Protoporphyrin IX, kommt oft lokalisiert in Flecken entweder direkt an der Oberfläche oder in eng-begrenzten Schichten innerhalb der Eierschale vor. Erbliche, pigmentbedingte Fleckung ist mit Faktoren wie Bruterfolg und Schalenstärke in Verbindung gebracht worden. In der vorliegenden Studie untersuchen wir unter standardisierten Bedingungen am Beispiel der Eierschalen zweier beliebter Studienobjekte unter den Singvögeln, der Kohlmeise (*Parus major*) und der Blaumeise (*Cyanistes caeruleus*), inwiefern die optische Kategorisierung von Eierschalenmustern wiederholbar ist, sowohl durch ein und denselben Beobachter, als auch zwischen verschiedenen

Beobachtern. Unsere Untersuchung zeigt für beide Meisenarten, dass die Wiederholbarkeit der Kategorisierung von Eierschalen beschränkt ist, sowohl im Vergleich wiederholter Kategorisierungen durch denselben Beobachter, als auch im Vergleich zwischen Beobachtern. Wir empfehlen daher in zukünftigen Untersuchungen alternative Methoden zur Quantifizierung von Fleckenmustern zu verwenden, zum Beispiel die digitale Bildanalyse, eine schon heute weit verbreitete Technik.

Introduction

Research into eggshell appearance and maculation (i.e. spotting) has been the focus of considerable ornithological studies in recent years (for full reviews, see Underwood and Sealy 2002; Kilner 2006; Reynolds et al. 2009; Cherry and Gosler 2010; Maurer et al. 2011). There are two key pigments responsible for the visible colouring and patterning of bird eggshells (Gorchein et al. 2009): protoporphyrin IX, responsible for red-brownish hues, and biliverdin, responsible for blue-green hues (Kennedy and Vevers 1973, 1976). Protoporphyrin is often localised as spots, either on the surface or in distinct layers within the eggshell (Baird et al. 1975; Kennedy and Vevers 1976).

Visual scoring methods of eggshell appearance were originally developed in the poultry sciences and used to compare eggshell colour with prepared standards composed from a series of "normal" eggs with shells of varying shades from white to dark brown (Wells 1968). Subsequently, eggshell pigment scoring has been applied to studies of wild bird species for recording the presence and coverage of maculation (see Kilner 2006) and quantifying the degree of this maculation (Gosler et al. 2000, 2005). The latter, more detailed pigment scoring method has been used in a variety of studies to compare maculation, particularly on the eggs of Great Tits (*Parus major*) and Blue Tits (*Cyanistes caeruleus*). In these two species, maculation has been scored in relation to factors such as breeding performance (e.g. Sanz and García-Navas 2009), eggshell strength (e.g. Gosler et al. 2005; Mägi et al. 2012), the laying female's ability to cope with anaemia (De Coster et al. 2012), and to investigate the inheritance (Gosler et al. 2000) and intra-clutch variation (De Coster et al. 2013) of eggshell patterning. This method has further been applied to the eggs of other species including House Sparrows (*Passer domesticus*) (López de Hierro and De Neve 2010), and Northern Lapwings (*Vanellus vanellus*) (Bulla et al. 2012). Although many of the afore-mentioned studies provided repeatability measures of their scores, none of them employed independent confirmation of the accuracy of this method.

Alternative methods for quantifying spot patterns exist, such as digital image analysis (e.g. Stoddard and Stevens 2010; Cassey et al. 2012). However, it has been shown that, under many circumstances, humans are better than computers at recognising and quantifying complex patterns (Kersten and Yuille 2003). Our brains are particularly specialised for interpreting variability in natural images, and this disguises the difficulty of dealing with the sophistication and ambiguity in complex patterns (Kersten et al. 2004). Therefore, eggshell pigment scoring by human visual assessment can potentially be a quick, cheap and reliable method to obtain information on eggshell spot patterns. In this study, we investigated whether pigment scoring of eggshell patterns is repeatable within and between observers by testing observers under standardised conditions. We used the eggshells of two passerines, Great Tits and Blue Tits, commonly used in such studies applying the spot scoring methods.

Methods

Eggshell sampling

Eggs were removed on the day of laying from a select number of clutches under license (Natural England Permit 20100857) as part of a different study (see Brulez et al. 2014). The study was conducted using an established box-nesting population of Great Tits and Blue Tits in Chaddesley Woods Nature Reserve, Worcestershire, UK (UK Ordnance Survey Grid Reference: SO914736, 52°36′N, 2°14′W). Following Cassey et al. (2010), eggs were photographed using a Canon 450D digital camera with a 105-MM Sigma AF lens under standardised conditions in the laboratory. The camera was raised on a Kaiser camera stand, surrounded by two Calumet photographic umbrellas with silver-white (AU3046) and flat white (AU3045) lining. The eggs were lit using two Osram 11-W energy saving light bulbs to the right and front. Photos were taken at ISO 400 with an aperture of f16 and the exposure was set to automatic. To ensure that the whole eggshell was recorded, four photographs were taken per eggshell, rotating the egg approximately 90° between photographs (Fig. 1).

Spot scoring

Eggshell maculation was recorded from photographic images. Following Gosler et al. (2000), eggshell pigmentation pattern was scored on the basis of three categories: pigment intensity [scored in 0.5 increments, from 0.5 (palest) to 5 (darkest)]; distribution [scored in 0.5 increments from 0.5 (>90 % of spots concentrated at a single

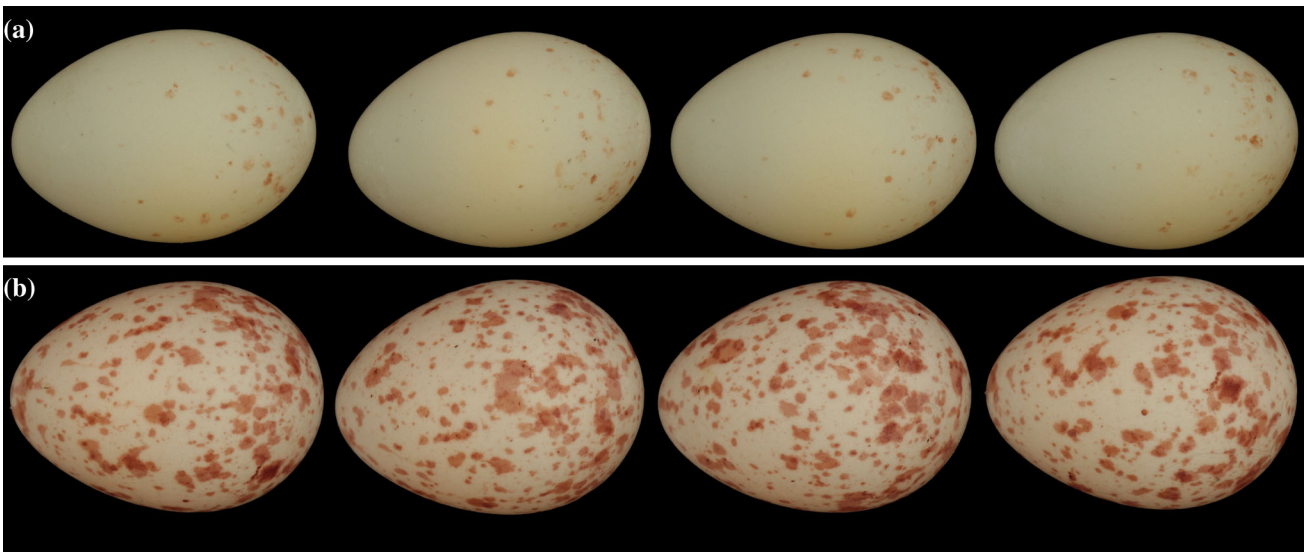

Fig. 1 An example of four photographs of the same **a** Blue Tit (*Cyanistes caeruleus*), and **b** Great Tit (*Parus major*) egg scored by each observers. Eggs were rotated approximately 90° between photographs to ensure that the whole eggshell was recorded

end) to 5 (spots evenly distributed)]; and spot size [scored in 0.5 increments from 1 (small spots) to 3 (large spots)]. For all three categories, 0 was scored for eggshells which were unspotted. This resulted in 11 possible scores for pigment intensity and distribution, and 7 possible scores for spot size.

All five observers were experienced ornithologists with similar knowledge and experience of eggshells but had never used the spot-scoring method prior to this study. As humans have personal variations in understanding and scoring natural patterns (de la Fuente de Val et al. 2006), guides to the scoring system (Fig. 1 in Gosler et al. 2000) were provided as a high-quality print out for simultaneous comparison. Prior to the onset of testing, each observer completed a training session, consisting of 20 randomised photographs not used in subsequent sessions. At the completion of the training session, the observers had the opportunity to qualitatively compare their scores with the photographs and the scores of the other observers presented in a spreadsheet.

Each observer scored a total of 408 (Great Tit = 184; Blue Tit = 224) photographs, consisting of 4 photographs per egg (Fig. 1), 2 eggs per clutch, from 51 clutches (Great Tit = 23; Blue Tit = 28). All viewing sessions were conducted on the same ViewSonic (VA2413wm) 24″ display screen with a Nvidia GeForce 7300 GT 512 Mb graphics processing unit. Testing consisted of ten sessions, with each session containing between 36 and 45 photographs. No session ever contained both Great Tit and Blue Tit eggs. Each observer was allowed 20 min to conduct a session, with at least a 20-min break between two consecutive sessions, and no more than four sessions on any one day. All observers conducted their sessions within five consecutive days. For each observer, the viewing order for all photographs was randomly assigned, and every image in every session was attributed a unique random four-character id (randomisation procedure written in SAS v.9.02 Proc IML).

Our design allowed us to compare inter-observer repeatability for all eggs and intra-observer repeatability for the same egg (4 photographs per egg; Fig. 1). To test the robustness of the results, we also compared the repeatability of scores for two eggs from the same clutch within observers. We hypothesised that scores would be more repeatable within observers than between observers, but that both scores would be more repeatable than scores on different eggs from the same clutch. In order to ascertain repeatability, we chose a threshold for the difference in two scores that we deemed biologically acceptable. Motivated by Gosler et al. (2000), we determined this threshold to be 0.5. We recognise that this threshold is somewhat arbitrary but we deemed it acceptable as it represents the smallest possible increment in scores, but is nevertheless a large proportion of the maximum values 3.0 (i.e. spot size) or 5.0 (i.e. spot intensity, distribution). We see no clearer justification for a different threshold. This threshold is used in conjunction with the tolerance interval approach described in Choudhary (2008) to ascertain repeatability. Essentially, this approach considers a scoring method to be repeatable if at least 80 % of differences in its scores fall within the threshold of acceptable differences.

Statistical analysis

We analysed the data in three steps. In the first step, we modelled the data using a variance components (or a

mixed-effects) model (Demidenko 2004). In the second step, we used the model fitted in the first step to define appropriate differences whose distributions contain information about the desired inter- and intra-observer repeatability. In the third step, we constructed tolerance intervals for these distributions and compared them with margin of acceptable differences to ascertain repeatability. This approach is common in concordance studies that compare multiple measurement methods (Choudhary 2008). It has previously been applied in oological studies to compare measurements of eggshell thickness using either micrometers or scanning electron microscopy (Igic et al. 2010). For further information on the statistical analyses, please refer to supplementary material 1.

Tolerance intervals $[-U, U]$ of the distributions of appropriately defined differences in two scores were used as a measure of repeatability, where U is the 95 % upper confidence bound of the *total deviation index* (Bland and Altman 1999; Lin 2000; Choudhary 2008). Having chosen $P = 0.80$, the tolerance interval provides an estimate of the range in which at least 80 % of differences in two scores are expected to lie with 95 % confidence. The confidence bound U, and hence the tolerance interval $[-U, U]$, was computed using the methodology described in Choudhary (2008). Specifically, we compute the following three intervals:

1. $[-U_1, U_1]$ to determine *inter-observer repeatability* for any two observers based on their different scores (five observers) for the same photo;
2. $[-U_2, U_2]$ to determine *intra-observer repeatability* for any single observer based on their multiple scores (four replicate photos) for the same egg; and
3. $[-U_3, U_3]$ to determine *intra-clutch repeatability* for any single observer based on their different scores for two eggs from the same female in a clutch.

Results

None of the upper bounds were found to be less than the chosen threshold (0.5), suggesting that spot scores are not repeatable (1) between observers, (2) within observers for photographs of the same egg, or (3) eggs from the same clutch. Table 1 presents the 95 % upper confidence bounds, U_1, U_2, and U_3, of the total deviation index ($P = 0.80$) that provide the three aforementioned tolerance intervals. The standard deviations of differences associated with these intervals ($\hat{\tau}_1$, $\hat{\tau}_2$ *and* $\hat{\tau}_3$) are also provided.

As an example, we describe one set of results in detail: the interval $[-U_1, U_1]$ for *inter-observer repeatability* for any two observers based on their different scores (five observers) for the same photo. The 95 % upper confidence bound for inter-observer agreement of scores of spot intensity is 1.34. This means that $[-1.34, 1.34]$ is a tolerance interval that captures at least 80 % of differences in scores of two observers with 95 % confidence. Scoring for spot intensity ranges from 0 to 5, so a difference of 1.34 is 26.8 % of the total range. The confidence bound for inter-observer agreement of scores of spot density is 1.85, 37 % of the total range (0–5). Similarly, the confidence bound for inter-observer agreement of scores of spot size is 0.73, 24.3 % of the total range (0–3). Photographic examples of eggshells whose differences in scores either fell, or did not fall, within the threshold of 0.5 for within observer differences and those for between observer differences are shown in Fig. 2.

Discussion

The human visual system is specialised for interpreting natural images and is the most complex pattern recognition device known (Kersten and Yuille 2003). We found that eggshell scores were neither repeatable within nor between

Table 1 Estimates of the standard deviations of the differences ($\hat{\tau}$) and associated 95 % upper confidence bounds (U) of the total deviation index with $P = 0.80$

	Two scorers looking at the same photo (inter-observer)		Two photos of the same egg (intra-observer)		Two eggs from the same clutch (intra-observer)	
	$\hat{\tau}_1$	U_1	$\hat{\tau}_2$	U_2	$\hat{\tau}_3$	U_3
Blue Tit (*Cyanistes caeruleus*)						
Intensity	0.86	1.34	0.70	0.94	1.10	1.61
Density	1.16	1.85	0.97	1.29	1.16	1.63
Size	0.50	0.73	0.48	0.64	0.59	0.83
Great Tit (*Parus major*)						
Intensity	0.86	1.32	0.68	0.91	0.75	1.02
Density	1.21	1.91	1.04	1.40	1.07	1.44
Size	0.55	0.78	0.51	0.69	0.62	0.87

The estimated standard deviations and confidence bounds determine inter- and intra-observer repeatability of scores gathered by the spot scoring method to record eggshell maculation in two species of passerine, Blue Tit and Great Tit

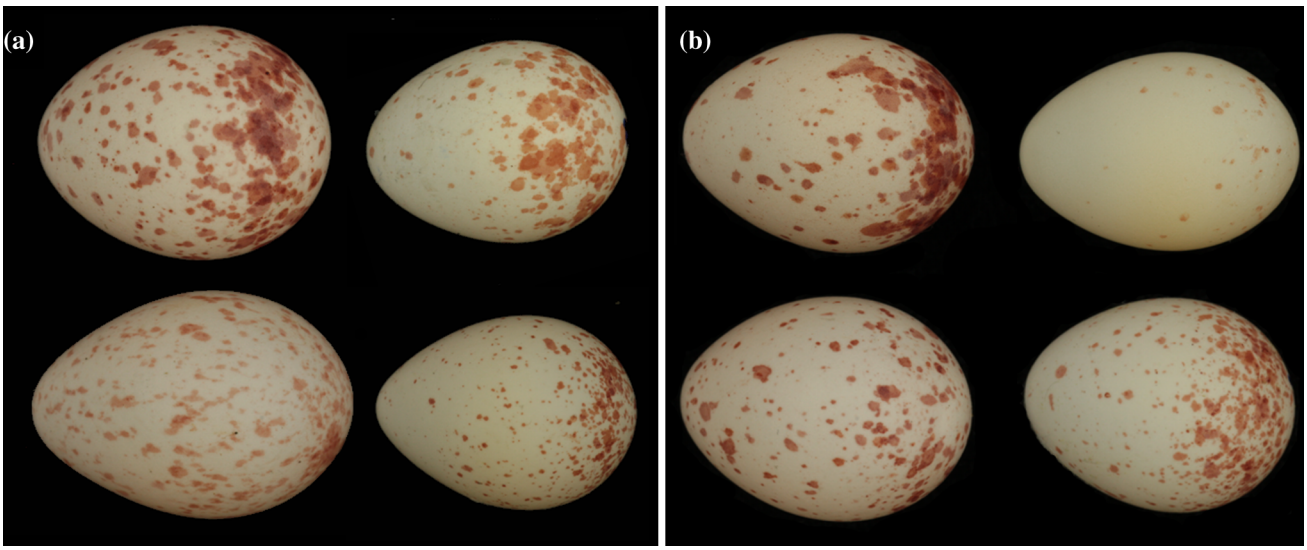

Fig. 2 **a** An example of a Great Tit (*left panel*) and a Blue Tit (*right panel*) eggshell which received scores from two observers (inter-observer) that differed by less than the threshold 0.5 (*top row*), compared with scores that differed by more than the threshold 0.5 (*bottom row*). **b** An example of a Great Tit (*left panel*) and a Blue Tit (*right panel*) eggshell which received repeated scores by any single observer (intra-observer) that differed by less than the threshold 0.5 (*top row*), compared with scores that differed by more than the threshold 0.5 (*bottom row*). Repeatable eggshells (i.e. the *top row*, which differed by less than the threshold 0.5) were deemed repeatable for all three categories of eggshell spotting (i.e. spot intensity, density, and size); however, unrepeatable spot scores (i.e. *bottom row*, which differed by more than the threshold 0.5) were only consistently unrepeatable for spot intensity and distribution, but not always for spot size

observers for either species, and, consequently, as expected, different eggs from the same clutch also did not produce repeatable scores.

Some eggshell spotting traits were more repeatable than others. Spot scores for size varied only slightly more than the required threshold of 0.5, representing the smallest possible increment in scores. However, due to the categorical set-up of the scoring method, this variable has fewer possible scores (range of possible scores = 7) than either spot intensity or distribution (range of possible scores = 11). It is more likely, therefore, to be repeatable by design. Scores for spot intensity were also close to being repeatable; however, only within observers. This stresses the importance of clearly defining score units prior to applying the method, as observers easily have different interpretations of how to apply the scores. Although spot score guides were provided during the scoring sessions, humans have personal variations in understanding and scoring natural patterns (de la Fuente de Val et al. 2006). Observing the scores for individual eggs (Fig. 2), it is clear that eggshells with spots receiving scores at the extreme ends (i.e. eggshells receiving scores 0.5 and 5.0 for pigment intensity and distribution and scores of 0.5 and 3.0 for spot size) were deemed to be more repeatable.

Pigment scores quoting significant intra- and/or inter-observer repeatability have been used in a variety of studies (e.g. Sanz and García-Navas 2009; Holveck et al. 2012). It is possible that our results differ from these studies for one or all of several reasons. Firstly, in our study, eggshells were scored retrospectively from photographs, taken under standardised lighting, on a computer. The full extent (i.e. variability) of spotting may have, therefore, been greater. Secondly, unlike in situ experiments where spot scoring would most likely be conducted opportunistically throughout the breeding season, in our study all eggshells were scored in 20-min sessions of >30 eggs. Although unlikely, it is possible that observers became fatigued or perturbed. Thirdly, observers may have suspected the purpose of the study was to perform 'well' and may have felt pressure to provide repeatable scores and were unduly influenced by previous sessions. Nevertheless, our results bring into question previous findings relating visually-scored maculation to factors such as eggshell strength (e.g. Gosler et al. 2005, 2011), clutch quality (e.g. Sanz and García-Navas 2009) and female quality (De Coster et al. 2013).

As some of the scores, particularly spot size, were close to the accepting threshold, it might be possible that with sufficient training, observers could produce repeatable scores. Familiarisation with a new method of measurement decreases measurement error (e.g. Yezerinac et al. 1992) and, as observers gain more experience, observer effects are known to decrease (e.g. Herring et al. 1994). Nevertheless, we found that among a group of highly trained ornithologists, eggshell spot-scoring was unsatisfactorily repeatable. We encourage future studies to use alternative methods for quantifying spot patterns, such as digital image analysis, which have already been extensively used (e.g.

Stoddard and Stevens 2010; Cassey et al. 2012). Elsewhere, it has been shown that the quantification of visible pigment spots cannot be used as a reliable proxy to determine protoporphyrin concentration of eggshells (Brulez et al. 2014), at least not in the two species of Tit studied here. For studies wishing to estimate the degree of protoporphyrin pigmentation we recommend directly quantifying the concentration of the pigment itself.

Acknowledgments This research was funded by the Natural Environment Research Council (NERC) through a studentship to K.B. The Worcestershire Wildlife Trust generously allowed the use of Chaddesley Woods National Nature Reserve. P.C. is an ARC Future Fellow (FT0991420). We are grateful to Franz Bairlein and two anonymous reviewers for comments that greatly improved the previous version.

References

Baird T, Solomon SE, Tedstone DR (1975) Localisation and characterisation of egg shell porphyrins in several avian species. Br Poult Sci 16(2):201–208. doi:10.1080/00071667508416177

Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8(2):135–160

Brulez K, Cassey P, Meeson A, Mikšík I, Webber SL, Gosler AG, Reynolds SJ (2014) Eggshell spot scoring methods cannot be used as a reliable proxy to determine pigment quantity? J Avian Biol 45(1): 94–102. doi:10.1111/j.1600-048X.2013.00236.x

Bulla M, Šálek M, Gosler AG (2012) Eggshell spotting does not predict male incubation but marks thinner areas of a shorebird's shells. Auk 129(1):26–35. doi:10.1525/auk.2012.11090

Cassey P, Portugal SJ, Maurer G, Ewen JG, Boulton RL, Hauber ME, Blackburn TM (2010) Variability in avian eggshell colour: a comparative study of museum eggshells. PLoS ONE 5(8):e12054. doi:10.1371/journal.pone.0012054

Cassey P, Thomas GH, Portugal SJ, Maurer G, Hauber ME, Grim T, Lovell PG, Mikšík I (2012) Why are birds' eggs colourful? Eggshell pigments co-vary with life-history and nesting ecology among British breeding non-passerine birds. Biol J Linn Soc 106(3):657–672. doi:10.1111/j.1095-8312.2012.01877.x

Cherry MI, Gosler AG (2010) Avian eggshell coloration: new perspectives on adaptive explanations. Biol J Linn Soc 100(4):753–762. doi:10.1111/j.1095-8312.2010.01457.x

Choudhary PK (2008) A tolerance interval approach for assessment of agreement in method comparison studies with repeated measurements. J Stat Plan Infer 138(4):1102–1115

De Coster G, De Neve L, Lens L (2012) Intraclutch variation in avian eggshell pigmentation: the anaemia hypothesis. Oecologia 170(2):297–304. doi:10.1007/s00442-012-2304-1

De Coster G, De Neve L, Lens L (2013) Intra-clutch variation in avian eggshell pigmentation covaries with female quality. J Ornithol. doi:10.1007/s10336-013-0974-z

de la Fuente de Val G, Atauri JA, de Lucio JV (2006) Relationship between landscape visual attributes and spatial pattern indices: a test study in Mediterranean-climate landscapes. Landsc Urban Plan 77(4):393–407

Demidenko E (2004) Mixed models: theory and applications. Wiley, New York

Gorchein A, Lim CK, Cassey P (2009) Extraction and analysis of colourful eggshell pigments using HPLC and HPLC/electrospray ionization tandem mass spectrometry. Biomed Chromatogr 23(6):602–606. doi:10.1002/bmc.1158

Gosler AG, Barnett PR, Reynolds SJ (2000) Inheritance and variation in eggshell patterning in the great tit *Parus major*. Proc R Soc Lond B 267(1461):2469–2473. doi:10.1098/rspb.2000.1307

Gosler AG, Higham JP, Reynolds SJ (2005) Why are birds' eggs speckled? Ecol Lett 8(10):1105–1113. doi:10.1111/j.1461-0248.2005.00816.x

Gosler AG, Connor OR, Bonser RHC (2011) Protoporphyrin and eggshell strength: preliminary findings from a passerine bird. Avian Biol Res 4(4):214–223. doi:10.3184/175815511X13207833399666

Herring W, Miller D, Bertrand J, Benyshek L (1994) Evaluation of machine, technician, and interpreter effects on ultrasonic measures of backfat and longissimus muscle area in beef cattle. J Anim Sci 72(9):2216–2226

Holveck M-J, Grégoire A, Staszewski V, Guerreiro R, Perret P, Boulinier T, Doutrelant C (2012) Eggshell spottiness reflects maternally transferred antibodies in blue tits. PLoS ONE 7(11):e50389

Igic B, Hauber ME, Galbraith JA, Grim T, Dearborn DC, Brennan PL, Moskát C, Choudhary PK, Cassey P (2010) Comparison of micrometer-and scanning electron microscope-based measurements of avian eggshell thickness. J Field Ornithol 81(4):402–410

Kennedy GY, Vevers HG (1973) Eggshell pigments of the araucano fowl. Comp Biochem Physiol B 44(1):11–25

Kennedy GY, Vevers HG (1976) Survey of avian eggshell pigments. Comp Biochem Physiol B 55(1):117–123. doi:10.1016/0305-0491(76)90183-8

Kersten D, Yuille A (2003) Bayesian models of object perception. Curr Opin Neurobiol 13(2):150–158

Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Annu Rev Psychol 55:271–304

Kilner RM (2006) The evolution of egg colour and patterning in birds. Biol Rev 81(3):383–406. doi:10.1017/S1464793106007044

Lin LI (2000) Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. Stat Med 19(2):255–270

López de Hierro M, De Neve L (2010) Pigment limitation and female reproductive characteristics influence egg shell spottiness and ground colour variation in the house sparrow (*Passer domesticus*). J Ornithol 151(4):833–840

Mägi M, Mänd R, Konovalov A, Tilgar V, Reynolds S (2012) Testing the structural-function hypothesis of eggshell maculation in the Great Tit: an experimental approach. J Ornithol 153:645–652. doi:10.1007/s10336-011-0782-2

Maurer G, Portugal SJ, Cassey P (2011) Review: an embryo's eye view of avian eggshell pigmentation. J Avian Biol 42(6):494–504. doi:10.1111/j.1600-048X.2011.05368.x

Reynolds SJ, Martin GR, Cassey P (2009) Is sexual selection blurring the functional significance of eggshell coloration hypotheses? Anim Behav 78(1):209–215. doi:10.1016/j.anbehav.2009.03.003

Sanz JJ, García-Navas V (2009) Eggshell pigmentation pattern in relation to breeding performance of Blue tits *Cyanistes caeruleus*. J Anim Ecol 78(1):31–41. doi:10.1111/j.1365-2656.2008.01465.x

Stoddard MC, Stevens M (2010) Pattern mimicry of host eggs by the common cuckoo, as seen through a bird's eye. Proc R Soc Lond B 277(1686):1387–1393. doi:10.1098/rspb.2009.2018

Underwood TJ, Sealy SG (2002) Adaptive significance of egg colouration. Avian incubation, behaviour, environment and evolution. Oxford University Press, Oxford

Wells RG (1968) The measurement of certain egg quality characteristics: a review. In: Carter TC (ed) Egg quality: a study of the hen's egg. Oliver & Boyd, Edinburgh

Yezerinac SM, Lougheed SC, Handford P (1992) Measurement error and morphometric studies: statistical power and observer experience. Syst Biol 41(4):471–482. doi:10.2307/2992588