Animal Behaviour 106 (2015) 51–57

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

Reproducible research in the study of biological coloration

Thomas E. White [a, *, †], Rhiannon L. Dalrymple [b, †], Daniel W. A. Noble [a, b, †], James C. O'Hanlon [a, †], Daniel B. Zurek [c, †], Kate D. L. Umbers [d, e, f, †]

[a] Department of Biological Sciences, Macquarie University, North Ryde, NSW, Australia
[b] Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, NSW, Australia
[c] Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, U.S.A.
[d] School of Biological Sciences, University of Wollongong, Wollongong, NSW, Australia
[e] Centre for Evolutionary Biology, University of Western Australia, Crawley, WA, Australia
[f] School of Science and Health, University of Western Sydney Hawkesbury, NSW, Australia

ARTICLE INFO

Article history:
Received 21 December 2014
Initial acceptance 19 January 2015
Final acceptance 21 April 2015
Published online
MS. number: 14-01044R

Keywords:
methods
photography
pigment
spectrometry
structural colour
visual ecology
visual modelling

The study of colour in nature has generated insights into fundamental evolutionary and ecological processes, and research into colour traits is a rapidly growing field (Kelber & Osorio, 2010). The ongoing interest in biological coloration has in part been driven by the increased availability of key technologies, including spectrometry and photography, and concurrent advances in methods for analysing colour data, such as visual models (e.g. Endler & Mielke, 2005; Kelber, Vorobyev, & Osorio, 2003; Stevens, Parraga, Cuthill, Partridge, & Troscianko, 2007). While these developments are positive for the field, the increasingly complex analyses being run on ever greater amounts of data heighten the need for comprehensive methods reporting and diligent data management (Alsheikh-Ali, Qureshi, Al-Mallah, & Ioannidis, 2011; Nekrutenko & Taylor, 2012).

Replication and transparency lie at the heart of science. Beyond simply allowing independent verification of results, reproducible research ensures greater comparability between studies and provides a foundation for testing new ideas and methods (Piwowar, Day, & Fridsma, 2007; Van Noorden, 2011; Whitlock, 2011). A study may be considered truly reproducible when it satisfies three broad criteria: (1) methods are reported completely, (2) data are publicly available and archived, and (3) the chain of modification of raw data is documented and preserved. While completely reproducible research (e.g. FitzJohn et al., 2014) is a laudable goal, the considerable demands it imposes on researchers means that it will often, in practice, be unattainable. Nevertheless, even partial reproducibility through the relatively simple practices of complete methods reporting and public data archiving is of tremendous value.

Our aim was to explore the state of reproducibility in the study of biological coloration, and to suggest simple ways in which it may be improved. We first outline common methods for studying biological coloration and present guidelines for comprehensive methods reporting. We then explore how well some of these important criteria have been reported in the literature. We also quantify the availability of publicly archived data and code and suggest some useful tools for increasing the reproducibility of colour trait research more broadly.

* Correspondence: T. E. White, Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia.
E-mail address: thomas.white@mq.edu.au (T. E. White).
† Authors contributed equally and are presented in random order.

http://dx.doi.org/10.1016/j.anbehav.2015.05.007

MEASURING COLOUR

Generations of biologists have endeavoured to explain the mechanisms and functions of animal and plant coloration (e.g. Endler & Mielke, 2005; Poulton, 1890; Thayer & Thayer, 1909; Wallace, 1891), and uncovering best practices in measuring colour has been a great challenge. The direct measurement of reflectance and/or transmittance through spectrometry revolutionized the study of biological coloration (Dyck, 1966), and has been widely adopted as the standard (Andersson, Prager, Hill, & McGraw, 2006). Digital photography is increasingly being used to quantify colour (Stevens et al., 2007), as high-resolution cameras are inexpensive and allow for the simultaneous, rapid sampling of multiple colour patches (McKay, 2013).

Expansion in the availability of objective methods for the measurement of colour has been matched by advances in theory and analysis. In particular, the development of visual models has enabled researchers to move beyond quantitative comparisons of reflectance spectra and adopt potentially more biologically relevant perspectives when defining and testing hypotheses (Chittka, 1992; Endler & Mielke, 2005; Vorobyev & Osorio, 1998). Visual models typically attempt to describe the reception and early stage processing of chromatic and achromatic information as a function of an object's reflectance, the ambient illumination and a receiver's sensory system (e.g. Endler & Mielke, 2005; Vorobyev & Osorio, 1998). Although relatively easy to implement, visual models are built on multiple assumptions about the way in which stimuli are processed that can dramatically shape the results of a given analysis (e.g. Lind & Kelber, 2009; Pike, 2012).

GUIDELINES FOR METHODS REPORTING

Given that methodological variation may shape results in significant and unpredictable ways (see Tables 1 and 2, and references therein), the comprehensive reporting of methods is a simple and crucial step in ensuring research is reproducible. Accordingly, we developed a list of information about the capture (Table 1) and analysis (Table 2) of colour data that should ideally be reported. With regard to the measurement of colour, we focus on the two most frequently used methods: photography and spectrometry. Analytical techniques are diverse, mathematically complex and are being developed rapidly (Kelber & Osorio, 2010; Théry & Gomez, 2010). Such progress means that the need for a deep understanding of common methods can quickly outstrip the working knowledge of the average researcher. As a consequence, the subtle complexity of many analytical techniques can be overlooked by empiricists, leading to critical methodological information not being reported. Our guidelines for reporting the details of colour analyses (Table 2) thus cover two broad, common methods: colorimetric (or 'spectral') analyses and visual modelling.

While we wish to emphasize that these details are essential to ensuring full reproducibility of data capture and analysis, we recognize that space restrictions in the main text of manuscripts may preclude the incorporation of all these details. In such cases, we suggest that details be included in meta-data or in a supplementary file so that the necessary information is available to researchers. It is also the case that there is variability in the degree to which each parameter may affect results, and so we have aimed to provide a brief, qualitative outline of the potential effects that variation in each parameter may have on the data (Tables 1 and 2). These tables are not intended as a guide to the selection of methods, however, for which we refer readers to excellent recent reviews as well as the original publications (Kelber et al., 2003; Kemp et al., 2015; Montgomerie, Hill, & McGraw, 2006; Stevens et al., 2007; and references in Tables 1 and 2).

ASSESSING REPORTING AND REPRODUCIBILITY

To determine the current state of reproducibility in the field we assessed a sample of the literature against a set of our criteria (Tables 1 and 2), which we expected should be commonly reported based on our background reading. We searched papers from 2013 in 22 leading journals: *American Journal of Botany*, *The American Naturalist*, *Animal Behaviour*, *Behavioral Ecology*, *Behavioral Ecology and Sociobiology*, *Biological Journal of the Linnean Society*, *Biology Letters*, *Current Zoology*, *Ecology*, *Ecology and Evolution*, *Ecology Letters*, *Ethology*, *Evolution*, *Functional Ecology*, *Journal of Ecology*, *The Journal of Evolutionary Biology*, *The Journal of Experimental Biology*, *Naturwissenschaften*, *New Phytologist*, *Oikos*, *PLoS One*, *Proceedings of the Royal Society B: Biological Sciences*. On each of the journals' homepages, we used the Boolean phrase 'colour' or 'color' or 'spectra*' to search the title and/or abstract. Journals were haphazardly divided up between the authors to review the 216 papers returned from our search. On first pass, we excluded review papers, methodological papers, papers quantifying spatial (i.e. pattern) rather than chromatic properties of a colour patch and studies taking microspectrometric measurements of retinal absorbance. The final set of 60 papers included only those that used either a spectrometer or camera to quantify coloration. To reduce the risk of observer bias in our assessment, each paper included in the final set was read and reassessed by two further authors. Any discrepancies between assessment scores were discussed by the three authors that had read the article and resolved prior to analysis. We also recorded whether data (in either a 'raw' or 'processed' form) and/or any code were publicly available. The data along with our analysis script have been stored as a github repository (http://dx.doi.org/10.5281/zenodo.16949). We have kept the papers used in our data set anonymous as our aim was to explore the general question of reproducibility in the field.

METHODS REPORTING IN COLOUR STUDIES

Our literature survey suggests there is surprising inconsistency and incompleteness in commonly reported methodological details (Fig. 1). Most studies ($N = 51$) used a spectrometer to measure colour, yet integration times (20%) and probe sample geometry (49%) and distance (20%) were often not reported. Among studies that used photography ($N = 18$), 67% detailed the number of pixels averaged, although camera models were more frequently reported (89%). Light sources were detailed in 76% and 65% of spectrometer- and camera-based studies, respectively. With regard to data analysis, of the 35 studies that calculated colorimetric variables, 77% specifically defined their measure of brightness, hue and/or chroma. The receptor noise-limited model (Vorobyev et al., 1998) was commonly used among studies with visual modelling (11 of 22), although there was considerable variation in the detailing of the type of receptor noise used (45% reported), the type of quantum catch used (59%) or whether photoreceptor adaptation (43%) was modelled. In contrast, details of the species' visual system being modelled (95%), the background used (82%) and the modelled illuminant (74%) were more commonly reported.

While some of the figures reported above seem troubling, it is important to note that 38% of papers made reference to previous work for details on some or all methods. The referenced works may have comprehensively covered some of these criteria, but were often incomplete as well, or referenced yet another paper. To avoid 'decay' of methodological detail reporting over successive papers, we suggest reporting all details along with the current manuscript. That aside, the remaining 62% of studies did not reference previous work, and were missing potentially important methodological details.

Table 1
Information about the capture of colour data we suggest be reported

Method	Information to be reported	Reason	Further discussion
Photography			Akkaynak et al., 2014; Stevens et al., 2007
	Lighting conditions	The stability and intensity of light across wavebands may affect results to varying degrees, depending on a study's objective. For the recovery of spectral radiance (the product of spectral illumination and surface reflectance) under natural daylight, for example, repeated calibration (i.e. linearization and equalization) against the light source may be essential. In contrast, for the recovery of reflectance data, the light source may have little effect as long as it is sufficiently stable and intense across relevant wavebands	
	Colour standards	Standards allow for the recovery of parameters of interest under varying lighting conditions, yet may differ considerably in their quality and comprehensiveness. Single light and dark standards, for example, are inadequate when outputs require linearization and equalization. Such cases require a larger set (ideally >5) of calibrated grey standards. Similarly, if attempting to estimate a camera's sensitivities, a large set of chromatic standards of known reflectance may be required	Bergman & Beehner, 2008
	Camera model	Cameras vary in the details of their construction (e.g. pixel size and gap), and manufacturers often provide useful information that may be easily found given knowledge of a camera's make and model	
	Camera optics	A camera's optics (e.g. lenses and optical filters) may shape results in several ways. For one, the optics in part define its sensitivity as they may selectively filter the light reaching the sensors. This is particularly relevant when studying objects with a UV component, for example, as standard glass lenses strongly filter UV light. Light may also not be uniformly transmitted across the lens surface, which will over- and underrepresent the intensity of pixels depending on their location in an image. Lower quality optics may also introduce spherical and, of particular concern, chromatic aberration	Ray, 2002
	Resolution	Higher resolution images capture finer-scale detail. Depending on the object of interest, low pixel resolution (<5 megapixels) images may miss important information, although most modern consumer-grade cameras capture relatively high resolution images by default	
	Sensor spectral sensitivities	The precise sensitivities of a camera's sensors are only required in limited circumstances, such as when mapping from camera-dependent outputs (RGB) to the sensitivities of an animal's visual system. The broad sensitivity range of a camera, however, may be detailed by manufacturers and would be valuable to report, particularly when studying UV reflective stimuli. Any sensor modifications (such as the removal of hot-mirrors) should also be reported	Pike, 2011
	Exposure	Exposure is jointly determined by aperture, shutter speed and ISO speed, all of which we suggest be reported. Camera calibration (hence, recorded values) may vary significantly with exposure, especially aperture, which necessitates either close control of a camera's exposure settings, or repeated recalibration. Variation in recorded values may be exacerbated by variation in aperture, as larger apertures allow in more light from the edge of a lens where chromatic and spherical aberrations, if present, will be more pronounced. Slower shutter speeds may negatively affect accuracy if samples move or the light source changes (e.g. in daylight)	
	White balance	Automatic white balancing is ubiquitous, but it is implemented differently across cameras and will often produce data in which outputs (e.g. RGB values) are incorrectly weighted. It is particularly problematic when using compressed file formats (e.g. JPEG) as white balance cannot be subsequently adjusted	
	Output file format	Compressed file formats (e.g. JPEG) may introduce chromatic and spatial artefacts to images, whereas uncompressed formats (i.e. TIFF and RAW) typically do not. The loss of information may be especially pronounced when studying fine-scale spatial and/or chromatic detail, although when studying large objects or averaging values across large image areas, the error introduced by file compression may be negligible. If compressed file formats are used, or compression is subsequently used during processing, the level of compression should be reported when possible so that the degree of error introduced may, in part, be estimated	Bergman & Beehner, 2008
	Software for image processing	There may be variation between programs, and between versions of individual programs, in the implementation of methods (e.g. some programs have camera-specific profiles for interpreting file formats). This may also affect results in cases where software bugs are known or are subsequently discovered	
	Linearization of sensor outputs	The colour channel outputs of cameras are often not linearly related to intensity, which is a requirement for quantitative image analysis. Channel outputs may therefore require independent linearization, and there are several methods to achieve this	Garcia, Dyer, Greentree, Spring, & Wilksch, 2013
	Equalization of sensor outputs	Colour channel outputs may not be equal, as cameras often show a wavelength bias which can lead to the over- and underrepresentation of certain colours in an image. For most purposes this will need to be corrected through a simple equalization process, although there are circumstances, such as when values are being mapped directly to another colour space and the camera's sensor sensitivities are known, where equalization may be largely unnecessary	
	Pixel sampling	Along with image resolution, choice of pixel sampling will influence the level of spatial detail captured, and variation in the data. The number of pixels averaged when sampling an area, the number of pixels per unit area and the size of the sampled area should ideally be reported	
Spectrometry			Andersson et al., 2006; Endler, 1990; Johnsen, 2012
	Light source	Higher intensity light results in lower noise, although a light source with sufficient stability and intensity across relevant wavebands should produce comparable results. However, error may be introduced when using a variable source (e.g. daylight), when comparing results from a column and a diffuse light source, or when the distance from the light source is significant (as noise increases with distance)	
	Light source, object and collector geometry	The angular relationship between the probe(s) and focal object may dramatically affect the recorded spectra, particularly for iridescent, glossy and irregularly shaped objects. The distance between the light source and focal object will alter the intensity of the light source and, hence,	Boyaci, Doerschner, Snyder, & Maloney,

(*continued on next page*)

Table 1 (*continued*)

Method	Information to be reported	Reason	Further discussion
		accuracy (with higher intensity light resulting in less noise). Probe distances will also partly define the size of the area being sampled, so we suggest both distances and angles be reported	2006; Meadows, Morehouse, Rutowski, Douglas, & McGraw, 2011; Santos, De Neve, Lumeij, & Förschler, 2007
	White/dark standard	Reflectance is a relative measure, and is dependent on the white and dark standards used in a given study. If calibrated standards are used, then results should be accurate and invariant. Uncalibrated standards, however, may introduce error	
	Spectrometer type and model	Spectrometers vary in their spectral sensitivity, wavelength resolution, and method (e.g. scanning, multichannel and hyperspectral). Manufacturers often provide useful information that may be easily found with knowledge of an instrument's make and model	
	Fibre diameter	The aperture of modern spectrometers is defined by the cross-sectional area of the fibre. Larger fibres thus increase sensitivity at the cost of decreased resolution. For many broadly reflecting natural spectra the results will be minimally affected by diameter variation, although considerable variation may arise when measuring objects with narrow spectral peaks, such as many structurally coloured animals	Akkaynak, 2014
	Software for spectral capture	There may be variation between programs, and between versions of individual programs, in the implementation of methods. This may affect results in cases where software bugs are known or are subsequently discovered	
	Integration time	This sets the period of time over which the light detected by the spectrometer's charge-coupled device (CCD) array is summed (analogous to a camera's shutter speed). Higher integration times improve sensitivity, but may negatively affect accuracy if samples move, the light source changes (e.g. in daylight), or if channels become saturated. If a specialized light source is used, as is typical in studies measuring the reflectance of biological materials, then intensity is generally not a concern and integration times may be short (e.g. ca. 50 ms)	
	Boxcar width	The boxcar width defines the number of adjacent pixels on the spectrometer's CCD array being averaged (i.e. the number of points in the moving average). Larger widths will improve accuracy through the reduction of noise, but will also reduce precision as spectral resolution is effectively lowered through channel averaging. Larger widths will also slightly increase sensitivity, although not as much as longer integration times. Error may be introduced if excessive smoothing is applied, although since natural spectra are generally smooth, modest values (ca. 10) tend to give sound results	
	Number of spectra averaged	Averaging spectral scans improves accuracy through the reduction of noise, and also slightly increases sensitivity	Dalrymple, Hui, Flores-Moreno, Kemp, & Moles, 2014
	Postcapture processing	Common postcapture manipulations of spectral data include smoothing, binning and the correction of negative values (e.g. through zeroing or the uniform addition of a minimum value). Each of these may affect both the accuracy and/or precision of measures, so all processing, along with relevant parameters (e.g. the value of α when LOESS smoothing), should ideally be reported	

The necessity of particular information will vary depending on a study's objectives, as will the potential amount of variation in results that may arise from variation in parameter choice. Nevertheless, we suggest these details be provided in the published paper or in meta-data to ensure that studies can be leveraged to their full potential.

Table 2
Information about the analysis of colour data we suggest be reported

Method	Information to be reported	Reason	Further discussion
	Reflectance spectra of stimuli	Figures of raw or aggregated spectra allow readers to visually inspect the nature and quality of spectral data, independent of further processing	
Colorimetric analyses			Montgomerie et al., 2006
	Definitions of spectral measures of hue, saturation and/or brightness used	There are many commonly used measures of hue (>3), brightness (>5) and chroma (>12) that vary in meaning and appropriateness	
Visual modelling			Kelber & Osorio, 2010; Kelber et al., 2003; Kemp et al., 2015; Wyszecki & Stiles, 1982
	Form of quantum catch (e.g. raw or log-transformed)	Quantum catches are often transformed to approximate the perceptual scaling of colour sensation. The log transformation of quantum catch values is common, as it ensures that differences in receptor stimulation are proportional to their magnitude in accordance with the Weber–Fechner law	Endler & Mielke, 2005
	Viewer sensitivity	Photoreceptor sensitivities partly define the spectral information available to a viewer at the earliest stage of visual processing. The function(s) should ideally be provided and the details for generating them described (e.g. visual pigment template, opsin lambda max(s), optical filtering). In simple cases (e.g. without filtering), the opsin lambda max values and modelling template used may suffice. Modelling results tend to be relatively robust to subtle variation in sensitivity inputs	Lind & Kelber, 2009
	Illuminant/irradiance spectrum	The illuminant defines the light available to be reflected by a stimulus, and thus partly defines the radiance spectrum arriving at a viewer's eye. The spectrum should be provided, not merely described, unless it exists as a well-defined standard in which case it should be indicated (e.g. D65 standard daylight)	
	Chromatic adaptation (i.e. von Kries transformation)	Chromatic adaptation refers to the independent normalization of receptor mechanisms to average viewing conditions. In many cases, the modelling of von Kries adaptation (typically assuming a uniform response to the background across photoreceptor types) will have no effect on results as it will simply shift the absolute position of points in a colour space. It may, however, influence results when analysing a given stimulus against different viewing backgrounds, or when incomplete adaptation is modelled	Kelber et al., 2003; Vorobyev, Brandt, Peitsch, Laughlin, & Menzel, 2001
	Background	If modelling chromatic adaptation, the background is typically what receptors are assumed to be adapted to. When asking questions of discriminability at threshold, the background is often taken as the scene from which a focal stimulus is being distinguished and will thus fundamentally shape results. The spectrum or spectra should be provided with raw data, not merely described (e.g. 'average vegetation background')	
	Noise type (quantum and/or neural)	Noise may be modelled as being independent of ambient light intensity (i.e. 'neural' noise only; often assumed when modelling in high-intensity light), or to also account for the random nature of photon absorption (i.e. 'quantum' or 'shot' noise; typically included when modelling in low light conditions). The effect of the type of noise chosen will thus in large part depend on the modelled light levels, both of which should be reported	Vorobyev & Osorio, 1998
	Relative receptor density	Noise decreases (hence, modelled discrimination ability in that wavelength range improves) with increased relative receptor density.	Lind & Kelber, 2009;
	Signal-to-noise ratio of receptor(s)	Defines the amount of noise in a given receptor and, hence, partly sets the threshold for the discrimination of two-coloured stimuli. Results are particularly sensitive to variation in the relative noise in receptor channels	Lind & Kelber, 2009; Vorobyev & Osorio, 1998
	Weber fraction in receptor(s)	Weber fractions can be calculated from signal-to-noise ratios and relative receptor densities, but we recommend reporting this too, given that they are often used in an unclear way	Vorobyev & Osorio, 1998

Guidelines are for two of the most common analytical approaches: colorimetrics (i.e. 'spectral' measures of hue, saturation and/or brightness, independent of any visual system) and visual modelling. Approaches to visual modelling are diverse, so the necessity of parameters may vary depending on the method used.

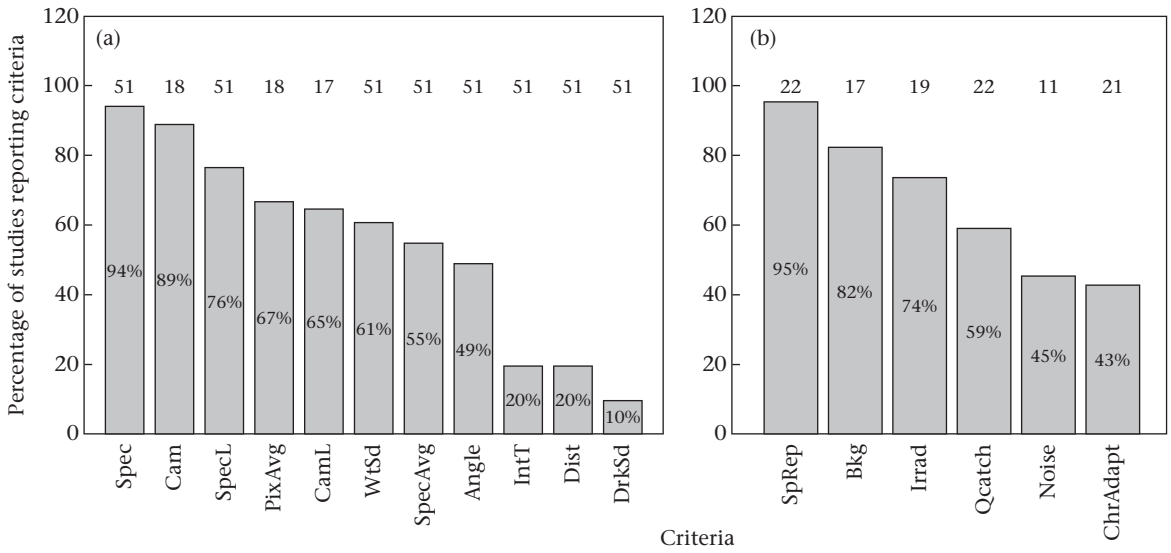

Figure 1. The comprehensiveness of methods reporting in the (a) measurement and (b) analysis of biological coloration assessed against a sample of more commonly reported information in Tables 1 and 2. Data are drawn from 60 empirical papers across 22 journals. Numbers above bars indicate sample size, as individual parameters may not be relevant to all studies. Spec: spectrometer model; Cam: camera model; SpecL: spectrometry light source; PixAvg: pixel averaging; CamL: photography light source; WtSd: white standard/s; SpecAvg: number of spectra averaged; Angle: spectrometry probe sample angle; IntT: spectrometry integration time; Dist: spectrometry probe sample distance; DrkStd: dark standard/s; SpRep: species of modelled receiver; Bkg: background; Irrad: illuminant; Qcatch: form of quantum catch; Noise: type of noise modelled; ChrAdapt: chromatic adaptation.

PUBLIC AVAILABILITY OF DATA AND CODE

Of the 60 studies analysed, 1.7% publicly provided the raw underlying data, 31.7% provided data in a preprocessed form and 66.7% of studies did not provide any publicly accessible data. The paucity of data being made available after publication in studies of biological coloration is in line with other fields (Drew et al., 2013; Vines et al., 2013; Wolkovich, Regetz, & O'Connor, 2012), including the broader field of animal behaviour (Caetano & Aisenberg, 2014). There is evidence, however, that this trend is shifting as funding agencies and publishers increasingly mandate the release of data (Whitlock, McPeek, Rausher, Rieseberg, & Moore, 2010).

The clearest benefit of open access data is that it allows researchers to build upon previous work for the testing and refinement of new ideas and methods. As discussed above, it is also essential that complete methodological details be provided, either in the manuscript or associated meta-data, to ensure data can be used to their full potential. This is particularly relevant in the study of colour, where new analysis methods are frequently developed (e.g. Allen & Higham, 2013; Endler, 2012; Stoddard, Kilner, & Town, 2014). More generally, the provision of open data is likely to foster collaboration as researchers draw upon existing work (Piwowar et al., 2007; Whitlock, 2011), and considerably reduce the difficulty of meta-analyses and large-scale comparative work (e.g. Burd, Stayton, Shrestha, & Dyer, 2014; Maia, Rubenstein, & Shawkey, 2013). We thus encourage researchers in our field to publicly archive raw data (e.g. individual reflectance spectra), to maximize its utility. The use of modest data embargoes and appropriate licences can help to ensure that original authors are able to make the best use of data and are subsequently credited, although we acknowledge that tensions over data sharing exist (discussed in Roche et al., 2014).

None of the included studies linked to any form of code. This is not surprising given the popularity of graphical user interface-based statistical and colour-analytical software (e.g. AVICOL; Gomez, 2006). When such software is freely available, and is combined with the comprehensive reporting of methods, its use represents a valuable approach to conducting reproducible research. We also note that there are advantages in adopting a programming workflow, in spite of the initial time investment required. Programming languages such as Python (Van Rossum & Drake, 1995) and R (R Core Team, 2014), for example, are free, open-source and are host to communities of developers that continuously build and implement analytical tools. Indeed, tools specifically for the analysis of colour data already exist (e.g. the package 'pavo' for R; Maia, Eliason, Bitton, Doucet, & Shawkey, 2013). While no program can compensate for having a clear understanding of the analysis at hand, programming languages are flexible by nature. This allows individual researchers to rapidly explore new methods as they are published, and methods-developers to implement their analyses to increase the reach and impact of their work. A further advantage of programmatic analyses is that analysis scripts, when properly curated, represent a complete documented history of a study's methods. The chain of modification of raw data may thus be preserved, both for researchers revisiting their own work in the future and for other scientists looking to build directly on the results of previous studies.

CONCLUSIONS

Our review of the recent literature highlights some impediments to reproducibility in the study of biological coloration. The simplest step towards reproducible research is through the comprehensive reporting of methods, yet key information is often unreported (Fig. 1). Here we provide a list of methodological information that we suggest be specified in studies that focus on the spectrometry- or camera-based analysis of colour (Tables 1 and 2). In addition to the suggested information, we also recommend that raw spectra be presented where possible, as these figures allow for the rapid assessment of the nature and quality of the data independent of downstream processing. We also emphasize the importance of explicitly outlining the biological justification underlying all choices for data capture and analysis (Kemp et al.,

2015). Finally, the public storage of data along with detailed metadata is pivotal for reproducibility, yet public data storage in studies of colour traits is not yet the norm. We therefore encourage researchers to consider the substantial benefits of publicly archiving their data (Piwowar et al., 2007; Whitlock, 2011). Overall, we hope that our guidelines will encourage researchers to think about the reproducibility of their work and the advantages of increased transparency, which will continue the advance of an exciting era in colour research.

Acknowledgments

We thank Rafael Maia, the Morehouse lab and three anonymous referees for valuable discussions and suggestions. T.E.W. thanks Elizabeth Mulvenna for endless support. T.E.W. and R.L.D. were supported by Australian Postgraduate Awards, and D.W.A.N. was supported by an NSERC PGSD and MQRES scholarship.

References

Akkaynak, D. (2014). Use of spectroscopy for assessment of color discrimination in animal vision. *Journal of the Optical Society of America A, 31*, A27–A33.
Akkaynak, D., Treibitz, T., Xiao, B., Gürkan, U. A., Allen, J. J., Demirci, U., et al. (2014). Use of commercial off-the-shelf digital cameras for scientific data acquisition and scene-specific color calibration. *Journal of the Optical Society of America A, 31*, 312–321.
Allen, W. L., & Higham, J. P. (2013). Analyzing visual signals as visual scenes. *American Journal of Primatology, 75*, 664–682.
Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. (2011). Public availability of published research data in high-impact journals. *PLoS One, 6*, e24357.
Andersson, S., Prager, M., Hill, G., & McGraw, K. (2006). Quantifying colors. In G. E. Hill, & K. J. McGraw (Eds.), *Bird coloration* (pp. 41–89). Cambridge, MA: Harvard University Press.
Bergman, T. J., & Beehner, J. C. (2008). A simple method for measuring colour in wild animals: validation and use on chest patch colour in geladas (*Theropithecus gelada*). *Biological Journal of the Linnean Society, 94*, 231–240.
Boyaci, H., Doerschner, K., Snyder, J. L., & Maloney, L. T. (2006). Surface color perception in three-dimensional scenes. *Visual Neuroscience, 23*, 311–321.
Burd, M., Stayton, C. T., Shrestha, M., & Dyer, A. G. (2014). Distinctive convergence in Australian floral colours seen through the eyes of Australian birds. *Proceedings of the Royal Society B: Biological Sciences, 281*, 20132862.
Caetano, D. S., & Aisenberg, A. (2014). Forgotten treasures: the fate of data in animal behaviour studies. *Animal Behaviour, 98*, 1–5.
Chittka, L. (1992). The colour hexagon: a chromaticity diagram based on photoreceptor excitations as a generalized representation of colour opponency. *Journal of Comparative Physiology A, 170*, 533–543.
Dalrymple, R. L., Hui, F. K., Flores-Moreno, H., Kemp, D. J., & Moles, A. T. (2014). Roses are red, violets are blue – so how much replication should you do? An assessment of variation in the colour of flowers and birds. *Biological Journal of the Linnean Society, 114*, 69–81.
Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., et al. (2013). Lost branches on the tree of life. *PLoS Biology, 11*, e1001636.
Dyck, J. (1966). Determination of plumage colours, feather pigments and-structures by means of reflection spectrophotometry. *Dansk Ornithologisk Forening, 60*, 49–76.
Endler, J. A. (1990). On the measurement and classification of colour in studies of animal colour patterns. *Biological Journal of the Linnean Society, 41*, 315–352.
Endler, J. A. (2012). A framework for analysing colour pattern geometry: adjacent colours. *Biological Journal of the Linnean Society, 107*, 233–253.
Endler, J. A., & Mielke, P. W. (2005). Comparing entire colour patterns as birds see them. *Biological Journal of the Linnean Society, 86*, 405–431.
FitzJohn, R. G., Pennell, M. W., Zanne, A. E., Stevens, P. F., Tank, D. C., & Cornwell, W. K. (2014). How much of the world is woody? *Journal of Ecology, 102*, 1266–1272.
Garcia, J. E., Dyer, A. G., Greentree, A. D., Spring, G., & Wilksch, P. A. (2013). Linearisation of RGB camera responses for quantitative image analysis of visible and UV photography: a comparison of two techniques. *PLoS One, 8*, e79534.
Gomez, D. (2006). *AVICOL: A program to analyse spectrometric data*. http://sites.google.com/site/avicolprogram/or from the author at dodogomezyahoo.fr.
Johnsen, S. (2012). *The optics of life: a biologist's guide to light in nature*. Princeton, NJ: Princeton University Press.
Kelber, A., & Osorio, D. (2010). From spectral information to animal colour vision: experiments and concepts. *Proceedings of the Royal Society B: Biological Sciences, 277*, 1617–1625.
Kelber, A., Vorobyev, M., & Osorio, D. (2003). Animal colour vision – behavioural tests and physiological concepts. *Biological Reviews, 78*, 81–118.
Kemp, D. J., Herberstein, M. E., Fleishman, L. J., Endler, J. A., Bennett, A. T. D., Dyer, A. G., et al. (2015). An integrative framework for the appraisal of colouration in nature. *American Naturalist*. http://dx.doi.org/10.1086/681021.
Lind, O., & Kelber, A. (2009). Avian colour vision: effects of variation in receptor sensitivity and noise data on model predictions as compared to behavioural results. *Vision Research, 49*, 1939–1947.
Maia, R., Eliason, C. M., Bitton, P.-P., Doucet, S. M., & Shawkey, M. D. (2013). pavo: an r package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution, 4*, 906–913.
Maia, R., Rubenstein, D. R., & Shawkey, M. D. (2013). Key ornamental innovations facilitate diversification in an avian radiation. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 10687–10692.
McKay, B. D. (2013). The use of digital photography in systematics. *Biological Journal of the Linnean Society, 110*, 1–13.
Meadows, M. G., Morehouse, N. I., Rutowski, R. L., Douglas, J. M., & McGraw, K. J. (2011). Quantifying iridescent coloration in animals: a method for improving repeatability. *Behavioral Ecology and Sociobiology, 65*, 1317–1327.
Montgomerie, R., Hill, G., & McGraw, K. (2006). Analyzing colors. In G. E. Hill, & K. J. McGraw (Eds.), *Bird coloration* (pp. 90–147). Cambridge, MA: Harvard University Press.
Nekrutenko, A., & Taylor, J. (2012). Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature Reviews Genetics, 13*, 667–672.
Pike, T. W. (2011). Using digital cameras to investigate animal colouration: estimating sensor sensitivity functions. *Behavioral Ecology and Sociobiology, 65*, 849–858.
Pike, T. W. (2012). Preserving perceptual distances in chromaticity diagrams. *Behavioral Ecology, 23*, 723–728.
Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One, 2*, e308.
Poulton, E. B. (1890). *The colours of animals: Their meaning and use especially considered in the case of insects*. London, U.K.: Kegan Paul, Trench, & Trübner.
R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Ray, S. (2002). *Applied photographic optics*. Oxford, U.K.: Focal Press.
Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., et al. (2014). Troubleshooting public data archiving: suggestions to increase participation. *PLoS Biology, 12*, e1001779.
Santos, S., De Neve, L., Lumeij, J., & Förschler, M. (2007). Strong effects of various incidence and observation angles on spectrometric assessment of plumage colouration in birds. *Behavioral Ecology and Sociobiology, 61*, 1499–1506.
Stevens, M., Parraga, C. A., Cuthill, I. C., Partridge, J. C., & Troscianko, T. S. (2007). Using digital photography to study animal coloration. *Biological Journal of the Linnean Society, 90*, 211–237.
Stoddard, M. C., Kilner, R. M., & Town, C. (2014). Pattern recognition algorithm reveals how birds evolve individual egg pattern signatures. *Nature Communications, 5*, 4117.
Thayer, G. H., & Thayer, A. H. (1909). *Concealing-coloration in the animal kingdom*. New York, NY: Macmillan.
Théry, M., & Gomez, D. (2010). Insect colours and visual appearance in the eyes of their predators. *Advances in Insect Physiology, 38*, 267–353.
Van Noorden, R. (2011). The trouble with retractions. *Nature, 478*, 26–28.
Van Rossum, G., & Drake, F. L., Jr. (1995). *Python reference manual*. Bloomington, IN: iUniverse Inc.
Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., et al. (2013). Mandated data archiving greatly improves access to research data. *FASEB Journal, 27*, 1304–1308.
Vorobyev, M., Brandt, R., Peitsch, D., Laughlin, S. B., & Menzel, R. (2001). Colour thresholds and receptor noise: behaviour and physiology compared. *Vision Research, 41*, 639–653.
Vorobyev, M., & Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society B: Biological Sciences, 265*, 351–358.
Wallace, A. R. (1891). *Natural selection and tropical nature: Essays on descriptive and theoretical biology*. New York, NY: Macmillan.
Whitlock, M. C. (2011). Data archiving in ecology and evolution: best practices. *Trends in Ecology & Evolution, 26*, 61–65.
Whitlock, M. C., McPeek, M. A., Rausher, M. D., Rieseberg, L., & Moore, A. J. (2010). Data archiving. *American Naturalist, 175*, 145–146.
Wolkovich, E. M., Regetz, J., & O'Connor, M. I. (2012). Advances in global change research require open science by individual researchers. *Global Change Biology, 18*, 2102–2110.
Wyszecki, G., & Stiles, W. S. (1982). *Color science*. New York, NY: Wiley.