

SCIENTIFIC REPORTS



OPEN

Species identification by experts and non-experts: comparing images from field guides

G. E. Austen¹, M. Bindemann², R. A. Griffiths¹ & D. L. Roberts¹

Received: 03 June 2016
Accepted: 31 August 2016
Published: 20 September 2016

Accurate species identification is fundamental when recording ecological data. However, the ability to correctly identify organisms visually is rarely questioned. We investigated how experts and non-experts compared in the identification of bumblebees, a group of insects of considerable conservation concern. Experts and non-experts were asked whether two concurrent bumblebee images depicted the same or two different species. Overall accuracy was below 60% and comparable for experts and non-experts. However, experts were more consistent in their answers when the same images were repeated, and more cautious in committing to a definitive answer. Our findings demonstrate the difficulty of correctly identifying bumblebees using images from field guides. Such error rates need to be accounted for when interpreting species data, whether or not they have been collected by experts. We suggest that investigation of how experts and non-experts make observations should be incorporated into study design, and could be used to improve training in species identification.

Accurate species identification is essential to ecological monitoring^{1,2}. Species observations are used to inform and evaluate conservation actions³, such as the monitoring of population trends^{4,5}, the implementation and evaluation of population management plans⁶, health assessments of ecosystems⁷, and extinction analysis⁸. Conversely, species misidentification can have serious negative impacts, such as the accidental culling of endangered species, exemplified by the threatened takahē *Porphyrio hochstetteri* (Meyer 1883) being mistaken for the destructive pukeko *Porphyrio porphyrio melanotus* (Temminck 1820)⁹, the incorrect monitoring of harmful algal blooms¹⁰, the unobserved decline in important fish stocks¹¹, and wasted resources, such as the drafting of inappropriate management plans from false species sightings¹².

While species identification in these contexts is conducted routinely by experts, such as taxonomists in museums or academic institutions¹³, there is also a long-standing tradition of members of the public supporting scientific research by contributing identification data¹⁴. Previously known as amateur naturalists, and more recently as citizen scientists, 70,000 of these lay recorders submit species observations on an annual basis in the UK alone¹⁵. These observers are recognised as a valuable asset in the monitoring of global environmental change^{16,17}. However, little is known about the accuracy of species identifications by non-experts, or how these compare to those of experts¹⁸. Although some doubts have been raised over the ability of volunteers to conduct 'real research'¹⁹, the assumption that recorded species have been correctly identified is rarely questioned¹. Yet the failure to account for possible species misidentification could affect assessments of population status and distribution and result in erroneous conservation decisions^{1,18,20}.

Few studies have investigated species identification accuracy. In a study of the classification of dinoflagellates, identification accuracy among expert observers was 72%¹⁰. Thus, more than one in four identifications was, in fact, a misidentification. Accuracy also varied dramatically, from 38% to 95%, depending upon the species being identified. However, accuracy was higher and more consistent in expert observers with field expertise than those with expertise gleaned from books¹⁰. In addition, individual consistency of experts with field expertise averaged 97% accuracy, but for those whose expertise came from books averaged only 75% accuracy. This indicates that observers with field experience were highly consistent in their decision-making (but both for correct and incorrect identifications), whereas the decisions of trained observers without such experience were more variable¹⁰.

A more recent study focused on the identification of individual mountain bongo antelopes *Tragelaphus eurycerus isaaci* (Thomas 1902) using a matching task²¹. In this task, expert and non-expert observers were shown

¹Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, Marlowe Building, University of Kent, Canterbury CT2 7NR, UK. ²School of Psychology, Keynes College, University of Kent, Canterbury CT2 7NP, UK. Correspondence and requests for materials should be addressed to G.E.A. (email: gea7@kent.ac.uk)

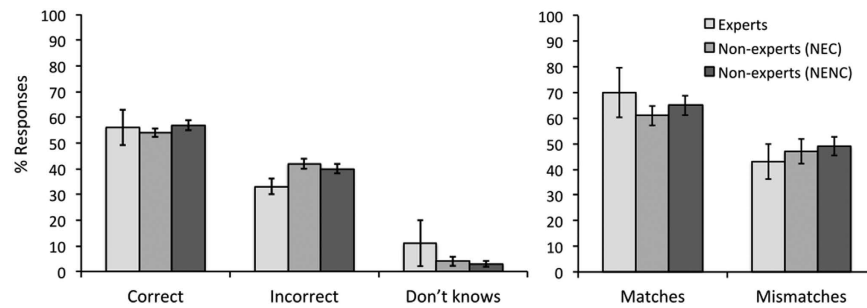


Figure 1. Percentage (± 1 s.e.) for correct, incorrect and “don’t know” responses (left graph), and accuracy (± 1 s.e.) for match and mismatch pairs (right graph) as a function of expertise. Overall accuracy is low and comparable (54% to 57%) between expert groups.

pairs of pictures of mountain bongos and had to decide whether these depicted the same or different individuals. Under these conditions, experts performed better than non-experts. However, accuracy was far from perfect in both groups of observers, with identification errors in at least one in five trials²¹.

These results suggest that observers can be prone to identification errors in species monitoring. However, whereas one study compared different types of experts during the identification of different species¹⁰, the other compared experts and non-experts during the identification of individuals from the same species²¹. Consequently, it is still unresolved how experts and non-experts compare directly in species identification.

In this study, we compare the identification accuracy of experts and non-experts with a matching task, in which observers have to decide whether pairs of images depict the same species. A key advantage of this task is that it allows for a direct comparison of observers with expertise in species identification with those without prior knowledge. This approach is used in other research areas, such as the study of forensic human face identification (see ref. 22), as an optimized scenario to establish best-possible performance²³, but little used in conservation research (although see ref. 21).

In order to investigate the accuracy of species identification in experts and non-experts, the model organisms should also be something that non-experts are familiar with. Bumblebees (*Bombus* sp.) are generally recognisable and attractive to members of the public²⁴, and are of great importance to human survival and the economy^{25–28}. Despite this importance, bee populations are in global decline from human activities^{26,29–32}. Consequently, bumblebees provide a relevant and timely model for studying the accuracy of species identification in expert and non-expert observers.

Experts and non-experts in bumblebee identification were asked to decide whether 20 pairs of bumblebee images depicted the same or two different species. To increase the relevance of this task to the monitoring of bumblebees by members of the public, the images used in this matching task were coloured illustrations of bumblebees taken from two easily accessible field guides. We sought to explore identification in detail by assessing the overall accuracy of observers in both groups, but also by exploring individual differences and the consistency of identification decisions. For this purpose, participants were asked to classify the same stimuli repeatedly, over three successive blocks.

Results

Participant expertise. Experts had substantial knowledge of bumblebees, whereas the non-experts knowledge was minimal. On average, experts named 20.7 bumblebee species (SD = 4.8; min = 15; max = 25), non-expert conservationists (NECs) named on average only 0.4 species (SD = 0.8; min = 0; max = 3), and non-expert non-conservationists (NENCs) only 0.2 species (SD = 0.4; min = 0; max = 1). Similarly, experts correctly chose an average of 19.7/20 UK species from a list of 40 *Bombus* species (Supplementary Table S2) (SD = 0.5; min = 19; max = 20), whereas NECs could only select an average of 1.6 species (SD = 2.1; min = 0; max = 7) and NENCs only 0.1 species (SD = 0.3; min = 0; max = 1).

Bee matching accuracy. Overall accuracy in the matching task was low and similar across groups of expertise (Fig. 1), with the mean percentage of correct responses ranging from 54% to 57%. Correspondingly, incorrect responses were high and recorded on between 33% (for experts) and 42% (NEC) of trials across groups. Finally, experts made “don’t know” responses on 11% of trials, while this contributed to less than 5% of responses in both groups of non-experts.

There was no difference between the three participant groups (E, NEC, and NENC) in terms of correct responses ($F(2,44) = 0.45$, $p = 0.638$), incorrect responses ($F(2,44) = 2.89$, $p = 0.066$), and “don’t know” responses ($F(2,44) = 0.35$, $p = 0.704$). Thus, experts and non-experts overall accuracy did not differ on this task. Match performance was similar across expertise groups, at between 43% and 49% accuracy, while mismatch performance was generally lower, at between 61% and 70% accuracy, but similar across the groups (Fig. 1). In line with these observations, a 3 (group: E, NEC, NENC) \times 2 (trial type: match, mismatch) mixed-factor ANOVA found an effect of trial type ($F(1,44) = 13.50$, $p = 0.001$), but not of expertise, ($F(2,44) = 0.61$, $p = 0.545$), and no interaction between factors ($F(2,44) = 0.59$, $p = 0.557$).

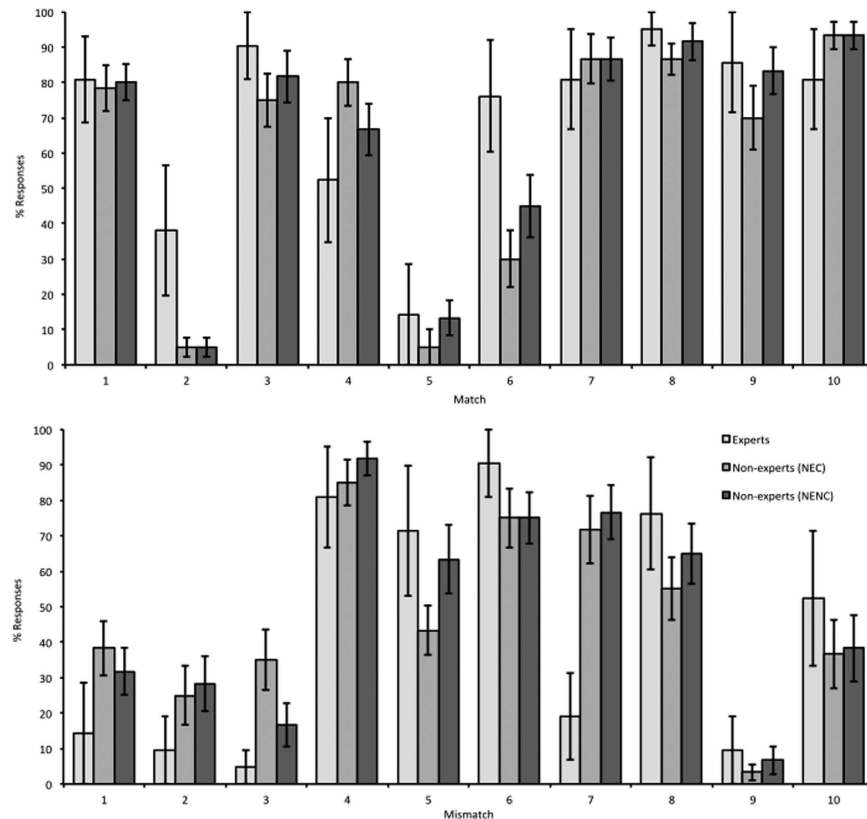


Figure 2. Mean accuracy (± 1 s.e.) across groups for each match (top) and mismatch (bottom) image. Effects of expertise were present for only three of these images (match 2, match 6 and mismatch 7).

Experience and matching accuracy. Percentage accuracy of correct, incorrect and “don’t know” responses were correlated with the years of experience that all participants reported in the identification of UK bumblebees. Correct and incorrect responses declined with experience ($r = -0.27$, $n = 47$, $p = 0.072$ and $r = -0.30$, $n = 47$, $p = 0.038$, respectively), but “don’t know” responses increased with experience ($r = 0.54$, $n = 47$, $p < 0.001$). This suggests that the more experienced observers were less likely to commit to a correct or incorrect identification decision. This inference is drawn tentatively, considering the limited sample of experts and possible extreme scores in the data (see *Participants* section).

Experience and accuracy for individual items. Accuracy was also calculated for all individual stimulus pairs and the groups of observers. For this by-item analysis, accuracy was combined across the three presentations of each stimulus pair (Fig. 2). One factor ANOVAs for each match and mismatch stimulus show that effects of expertise were present for only three of the 20 images. Post-hoc Tukey tests reveal that experts outperformed non-experts with Match 2 ($F(2,44) = 5.92$, $p = 0.005$; E v NEC and E v NENC $p = 0.007$) and Match 6 ($F(2,44) = 3.76$, E v NEC $p = 0.024$). Conversely, non-experts outperformed experts with Mismatch 7 ($F(2,44) = 7.00$, NEC v E $p = 0.005$ and NENC v E $p = 0.002$). This pattern suggests that the reliance on purely visual information by non-experts generally leads to comparable and occasionally even better accuracy than experts. This indicates that the additional subject-specific experience of experts does not consistently improve, and might even hinder, the matching of some bumblebee species. In some cases, however, expert knowledge also adds a performance advantage that must transcend the available visual information.

Consistency. We also sought to determine whether experts might be more consistent than non-experts in their identification of bumblebees, by assessing performance across the three repeated trials. Consistent decisions were defined as instances in which observers made the same responses to bumblebee pairs in all three trials.

Two consistency measures were obtained. The first of these reflects overall consistency regardless of accuracy, and was calculated by collapsing consistent correct (42% of all decisions), incorrect (27%), and don’t know (2%) responses for the different expertise groups (Fig. 3). A one-factor ANOVA showed that consistency varied between participant groups ($F(2,44) = 5.42$, $p = 0.008$). Tukey post-hoc test showed that experts were more consistent than the NEC ($p = 0.020$) and the NENC groups ($p = 0.006$). The consistency of NECs and NENCs did not differ ($p = 0.830$). A second consistency measure was calculated, which reflects the consistency of accurate responses only. This revealed a similar pattern, with experts outperforming the two non-expert groups. A one-factor ANOVA showed that these differences between groups were not reliable ($F(2,44) = 0.55$, $p = 0.583$). However, a correlation between consistent and consistently-accurate responses (Fig. 4) was found ($r = 0.722$,

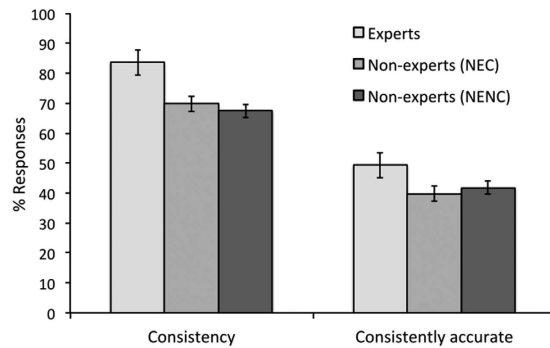


Figure 3. Percentage (± 1 s.e.) consistency in responses across presentations of stimuli. Experts were more consistent than non-experts in both overall and accurate answers.

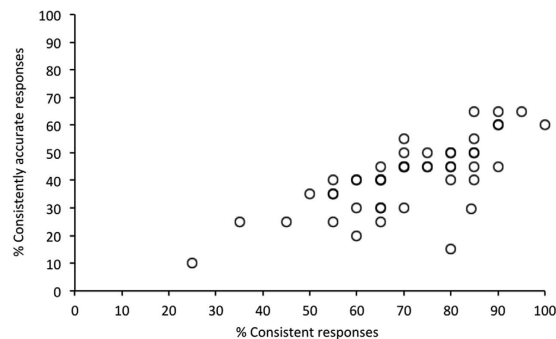


Figure 4. Correlation of consistent and consistently accurate responses. Individuals with consistent responses are more likely to be consistently accurate.

$n = 47$, $p < 0.001$). Taken together, these data indicate that experts are generally more consistent than non-experts in their responses, but not in their accurate responses. However, the individuals (experts or non-experts) whose responses are more consistent are also more likely to be consistently accurate.

Discussion

This study examined experts' and non-experts' accuracy in the identification of bumblebees. The naming tasks revealed clear effects of expertise, with experts naming an average of 20.7 UK bumblebee species and selecting 19.7/20 from a list of bumblebee species, while NECs and NENCs could name on average less than one species and selected less than two. A different picture emerged when the *actual* visual identification accuracy of these observers was assessed with the matching task. In this task, experts' overall accuracy was low, at 56%, and indistinguishable from NECs and NENCs. This finding was confirmed when performance was broken down into match and mismatch trials, for which expert and non-expert performance also did not differ. Participants' self-reported years of experience in bumblebee identification was also correlated with responses on the matching task. This analysis shows that both correct and incorrect responses decline with experience, but 'don't know' responses increase. Thus, observers appear to become more cautious with experience and less willing to commit to any identification decisions. These inferences are drawn tentatively, due to the limited availability of bumblebee experts for this study ($n = 7$). Crucially, however, these findings suggest once again that expertise does not improve the visual identification of bumblebees in the present task.

Overall, these findings converge with studies that have shown that visual species identification can be surprisingly error-prone. In contrast to previous studies, which either examined different types of experts during the identification of different species¹⁰, or experts and non-experts during the identification of individuals from the same species²¹, the current experiment compared experts and non-experts during species identification. The error rates that are observed across these studies raise important questions concerning the accuracy of species identification using field guides. Such identifications are used for supporting a wide range of actions, such as the monitoring of endangered species^{33–35} and the drafting of appropriate management plans^{36–38}. An understanding of error rates needs to be factored into such important conservation activities.

We draw these conclusions with some caveats. It is conceivable, for example, that accuracy among experts and non-experts will vary depending on how images are presented or which guidebooks are at hand. A number of identification guides exist for UK bumblebees, providing a variety of pictures of bumblebees, e.g. line or colour drawings, photographs and stylised diagrams. The extent to which illustrations from different guides accurately capture the key visual features of different bumblebee species and also match each other remains open to exploration³⁹, but is bound to affect bumblebee identification tasks. Variation in the specimens used by illustrators may also be due to phenotypic variation or even mislabelling in the collections used. Moreover, a difficult question for

illustrators is which individual of a species should be drawn in order for a guidebook to represent a ‘typical’ specimen. There is the option to use the holotype (the single specimen on which the species is described), but this may not be readily available or representative of the current UK population. There is the added complication of physical differences caused by age, such as the loss of hair and fading of colour due to the sun²⁴.

There is a long history of members of the public contributing to species monitoring programmes^{3,40,41}. An analysis of accuracy for individual items indicated that expert and non-expert accuracy was similar for most bumblebee species here. There were also instances in which experts outperformed non-experts or the reverse pattern was found. This indicates that for some species comparisons, the reliance on purely visual information (as available to both experts and non-experts) produces best accuracy. In other cases, the additional subject-specific expert knowledge can occasionally interfere with the visual identification process. However, expert knowledge can also transcend the available visual information in some cases and provide a benefit in performance. This mixture of results is an intriguing outcome that is perhaps counter to intuition, because it suggests that the identification accuracy of bumblebees might be optimized best by using expert and non-expert decisions in a complementary fashion.

Experts were more consistent in their decisions when the tests were repeated. This effect was only reliable when correct, incorrect and don’t know decisions were combined. Overall, however, the more consistent observers were also more consistently accurate. Thus, experts’ decision criteria appear to be more stable and this might confer an advantage when identification of the same species is assessed repeatedly. Further, systematic investigations of these different effects (visual vs. expertise-driven identifications, consistency) might inform training that is designed to enhance the accuracy of observers. This could help to reduce error rates and improve species monitoring in the field.

Additional identification cues might also be available that could specifically enhance expert performance in practical settings, such as context, behaviour, flight period, or even the presentation of a live or dead specimen. Experts will also have access to additional resources to support identification, such as taxonomic revisions with identification keys and diagrams or natural history collections. We included a short questionnaire in our study to assess field guide usage, which showed that all seven experts reported a combination of up to five field guides, and three also utilised smartphone apps. However, our data suggests that this experience did not enhance performance in the current experiment. More generally, it remains unresolved whether sufficient numbers of experts can be found for research to provide the volumes of data required to understand such factors⁴². Results may also differ for other taxa, but the growth of citizen science and the increase in use of these volunteer data means that species observations, such as those used to inform conservation practitioners, are likely to be heavily reliant on images, either as submissions by non-experts or validation by experts.

In conclusion, this study shows that experts and non-experts both make many errors when using standard field guide illustrations to identify species. This raises important questions surrounding the accuracy of species observations in ecological datasets, and suggests that consideration should be given to possible inaccuracies when such information is used to inform decision makers.

Methods

Participants. This research was approved by the Ethics Committee of the School of Psychology at the University of Kent (UKC) and conducted in accordance with the ethical guidelines of the British Psychological Association. Informed consent was obtained from all participants before taking part in the survey. A total of 47 people participated in the survey, comprising expert and non-expert observers. Seven experts (3 female, 4 male, mean age = 40 years, range 25–64) were recruited via a national non-governmental organisation (NGO) specialising in the conservation of bumblebees. Forty non-experts were recruited via the School of Anthropology and Conservation at UKC (30 female, 10 male, mean age = 35 years, range = 18–65). Half of these participants ($n = 20$; 15 female, mean age = 33 years) had a general background in nature conservation and were classified as non-expert conservationists (NEC). The remaining participants ($n = 20$; 15 female, mean age = 37 years) had little or no experience with nature conservation and were therefore classified as non-expert non-conservationists (NENC). All 47 participants reported good vision or corrected-to-normal.

The seven expert participants reported a total of 39 years experience (1–15 years) in the identification of bumblebee species, whereas only seven of the non-experts reported any experience in the identification of bumblebees, ranging from 1 to 8 years. To define this experience further, all participants were asked to evaluate their identification experience on a five-point scale. Self-evaluated bumblebee identification abilities of experts and non-experts did not overlap. Non-experts reported ‘no experience’ ($n = 33$), ‘little experience’ (6) and ‘some experience’ (1), while experts described themselves as ‘experienced’ (3) and ‘competent’ (4).

Stimuli. The stimuli consisted of 20 pairs of images of bumblebees, comprising 10 match pairs (same species shown), and 10 mismatch pairs (different species shown), using images from two different field guides^{43,44}. For match pairs, illustrations were from different artists (see Supplementary Table S1). Images in each pair consisted of colour illustrations of dorsal views of entire bumblebees, presented side-by-side on a white background. The paired images always displayed the same caste, e.g. both males, both queens etc. Stimuli were designed to be viewed on a computer monitor, and measured approximately 24 × 15 cm onscreen. No zoom function was included in the survey. Species names were taken from a checklist of extant, native bumblebees recorded in Britain and Northern Ireland (genus *Bombus* Latreille), downloaded from the Natural History Museum (London) website (www.nhm.ac.uk). For each of the species on this list, the BirdGuide application (‘app’) “Bumblebees of Britain and Ireland” (www.birdguides.com) was used to identify phenotypes associated with each species. For species that exhibited a different phenotype according to caste, an individual entry was listed for every caste that differed in appearance from other castes for each species in that guide. Although listed in the guide, *B. pomorum* (Panzer, 1805) and *B. cullumanus* (Kirby, 1802) are believed extinct, and so were removed. The randomised list

also included two species in the *lucorum* complex, *B. magnus* (Vogt, 1911) and *B. cryptarum* (Fabricius, 1775), but as research shows that these are visually inseparable²⁸, these were also removed. The final list comprised 45 entries representing different UK species and castes where applicable. Twenty entries were randomly sampled from the list for use in the tests. For the survey, the list of the 20 selected entries was randomised again, and the first 10 entries were the species and caste used to create match pairs. The remaining 10 entries on the list formed the first half of a mismatch pair, with the second half of the pair being selected from the other 19 species named on the list. The second species was chosen as so to create a mix of visually similar and dissimilar species (Supplementary Table S1). This set of 10 match and 10 mismatch images was used to create an online survey.

Procedure. In the experiment, participants' bumblebee knowledge was initially recorded using two simple tasks to assess their expertise. First, participants were asked to write down all bumblebee species found in the UK. Then participants were asked to select UK bumblebee species from a list of 40 bumblebees (20 UK and 20 non-UK species) (Supplementary Table S2). On completion of the initial assessments, participants were given the matching task. In this task, participants were asked to classify each pair of bumblebees as the same species, two different species, or provide a "don't know" response using three different buttons on a standard computer keyboard. No time limit was applied to this task to encourage best-possible performance. Participants completed three blocks of this task. Each of these comprised the 10 match and 10 mismatch pairs, and the order of presentation was randomised for the two repeats. In the experiment, each stimulus was therefore shown three times. Inferential statistics were performed using arcsine square-root transformed data.

References

1. Elphick, C. S. How you count counts: the importance of methods research in applied ecology. *J. Appl. Ecol.* **45**, 1313–1320 (2008).
2. Farnsworth, E. J. *et al.* Next-Generation Field Guides. *Bioscience* **63**, 891–899 (2013).
3. Sutherland, W. J., Roy, D. B. & Amano, T. An agenda for the future of biological recording for ecological monitoring and citizen science. *Biol. J. Linn. Soc.* **115**, 779–784 (2015).
4. Rodrigues, A. S. L., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M. & Brooks, T. M. The value of the IUCN Red List for conservation. *Trends Ecol. Evol.* **21**, 71–76 (2006).
5. Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M. & Elkinton, J. S. Observer bias and the detection of low-density populations. *Ecol. Appl.* **19**, 1673–79 (2009).
6. Duelli, P. Biodiversity evaluation in agricultural landscapes: An approach at two different scales. *Agric. Ecosyst. Environ.* **62**, 81–91 (1997).
7. Butchart, S. H. M. *et al.* Global biodiversity: indicators of recent declines. *Science* **328**, 1164–1168 (2010).
8. Roberts, D. L., Elphick, C. S. & Reed, J. M. Identifying anomalous reports of putatively extinct species and why it matters. *Conserv. Biol.* **24**, 189–196 (2010).
9. Hunt, E. New Zealand hunters apologise over accidental shooting of takahē. *The Guardian* (2015). Available at: <http://www.theguardian.com/environment/2015/aug/21/new-zealand-conservationists-apologise-over-accidental-shooting-of-endangered-takahē> (Accessed: 21st August 2015).
10. Culverhouse, P. F., Williams, R., Reguera, B., Herry, V. & González-Gil, S. Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Mar. Ecol. Prog. Ser.* **247**, 17–25 (2003).
11. Beerkircher, L., Arocha, F. & Barse, A. Effects of species misidentification on population assessment of overfished white marlin *Tetrapturus albidus* and roundscale spearfish *T. georgii*. *Endanger. Species Res.* **9**, 81–90 (2009).
12. Solow, A. *et al.* Uncertain sightings and the extinction of the ivory-billed woodpecker. *Conserv. Biol.* **26**, 180–184 (2012).
13. Hopkins, G. W. & Freckleton, R. P. Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Anim. Conserv.* **5**, 245–249 (2002).
14. Stepenuck, K. & Green, L. Individual- and community-level impacts of volunteer environmental monitoring: a synthesis of peer-reviewed literature. *Ecol. Soc.* **20** (2015).
15. Pocock, M. J. O., Roy, H. E., Preston, C. D. & Roy, D. B. The Biological Records Centre: a pioneer of citizen science. *Biol. J. Linn. Soc.* **115**, 475–493 (2015).
16. Johnson, M. F. *et al.* Network environmentalism: citizen scientists as agents for environmental advocacy. *Glob. Environ. Chang.* **29**, 235–245 (2014).
17. Sauermaun, H. & Franzoni, C. Crowd science user contribution patterns and their implications. *Proc. Natl. Acad. Sci.* **112**, 679–684 (2015).
18. Shea, C. P., Peterson, J. T., Wisniewski, J. M. & Johnson, N. A. Misidentification of freshwater mussel species (Bivalvia: Unionidae): contributing factors, management implications, and potential solutions. *J. North Am. Benthol. Soc.* **30**, 446–458 (2011).
19. Cohn, J. P. Citizen Science: Can volunteers do real research? *BioScience* **58**, 192–197 (2008).
20. Runge, J., Hines, J. & Nichols, J. Estimating species-specific survival and movement when species identification is uncertain. *Ecology* **88**, 282–287 (2007).
21. Gibbon, G. E. M., Bindemann, M. & Roberts, D. L. Factors affecting the identification of individual mountain bongo antelope. *PeerJ* **3**, e1303 (2015).
22. Johnston, R. A. & Bindemann, M. Introduction to forensic face matching. *Appl. Cogn. Psychol.* **27**, 697–699 (2013).
23. Burton, A. M., White, D. & McNeill, A. The Glasgow Face Matching Test. *Behav. Res. Methods* **42**, 286–291 (2010).
24. Edwards, M. & Jenner, M. *Field guide to the bumblebees of Great Britain & Ireland* (Ocelli, 2005).
25. Rains, G. C., Tomberlin, J. K. & Kulasiri, D. Using insect sniffing devices for detection. *Trends Biotechnol.* **26**, 288–294 (2008).
26. Potts, S. G. *et al.* Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* **25**, 345–353 (2010).
27. Klatt, B. K. *et al.* Bee pollination improves crop quality, shelf life and commercial value. *Proc. Biol. Sci.* **281**, 20132440 (2014).
28. Scriven, J. J. *et al.* Revealing the hidden niches of cryptic bumblebees in Great Britain: Implications for conservation. *Biol. Conserv.* **182**, 126–133 (2015).
29. Ghazoul, J. Buzziness as usual? Questioning the global pollination crisis. *Trends Ecol. Evol.* **20**, 367–373 (2005).
30. Gallai, N., Salles, J.-M., Settele, J. & Vaissière, B. E. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* **68**, 810–821 (2009).
31. Williams, P. H. & Osborne, J. L. Bumblebee vulnerability and conservation world-wide. *Apidologie* **40**, 367–387 (2009).
32. Cresswell, J. E. *et al.* Differential sensitivity of honey bees and bumble bees to a dietary insecticide (imidacloprid). *Zoology* **115**, 365–371 (2012).
33. Barlow, K. E. *et al.* Citizen science reveals trends in bat populations: The National Bat Monitoring Programme in Great Britain. *Biol. Conserv.* **182**, 14–26 (2015).
34. Dennhardt, A. J., Duerr, A. E., Brandes, D. & Katzner, T. E. Integrating citizen-science data with movement models to estimate the size of a migratory golden eagle population. *Biol. Conserv.* **184**, 68–78 (2015).

35. Lee, T. E. *et al.* Assessing uncertainty in sighting records: an example of the Barbary lion. *PeerJ* **3**, e1224 (2015).
36. Guisan, A. *et al.* Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**, 1424–1435 (2013).
37. Tulloch, A. I. T., Possingham, H. P., Joseph, L. N., Szabo, J. & Martin, T. G. Realising the full potential of citizen science monitoring programs. *Biological Conservation* (2013).
38. Lukyanenko, R., Parsons, J. & Wiersma, Y. Citizen science 2.0 : data management principles to harness the power of the crowd in *Service-Oriented Perspectives in Design Science Research* (eds Hemant, J., Sinha, A. P. & Vitharana, P.) 465–473 (Springer, 2011).
39. Fitzsimmons, J. M. How consistent are trait data between sources? A quantitative assessment. *Oikos* **122**, 1350–1356 (2013).
40. Bonney, R. *et al.* Citizen Science: A developing tool for expanding science knowledge and scientific literacy. *BioScience* **59**, 977–984 (2009).
41. Silvertown, J. A new dawn for citizen science. *Trends Ecol. Evol.* **24**, 467–471 (2009).
42. Kelling, S. *et al.* Taking a 'big data' approach to data quality in a citizen science project. *Ambio* **44**, 601–611 (2015).
43. Prŷs-Jones, O. E. & Corbet, S. A. *Bumblebees* (Richmond Publishing Company, 1991).
44. BBC Wildlife Pocket Guides Number 11 *Bumblebees other bees and wasps* (BBC Wildlife Magazine, 2006).

Acknowledgements

Many thanks to those that participated in this survey, and to those who circulated the survey. Special thanks go to the Bumblebee Conservation Trust (BBCT) who promoted the survey to experts. We are grateful to Dr Nikki Gammans from BBCT, Stuart Roberts from the Bees Wasps and Ants Recording Society (BWARS) and Dr Paul Williams from the Natural History Museum (NHM) for their input and advice given before designing the survey. This project is part of a PhD funded by a University of Kent 50th Anniversary Research Scholarship.

Author Contributions

G.E.A., M.B. and D.L.R. conceived the idea and designed the experiment. G.E.A. and M.B. wrote the main manuscript text and prepared all figures. G.E.A., M.B. and R.A.G. decided on which analyses to use and G.E.A. conducted the analyses. All authors edited the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Austen, G. E. *et al.* Species identification by experts and non-experts: comparing images from field guides. *Sci. Rep.* **6**, 33634; doi: 10.1038/srep33634 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016