Seeing rare birds where there are none: self-rated expertise predicts correct species identification, but also more false rarities

Nils Bouillard[1], Rachel L. White[2], Hazel A. Jackson[3], Gail E. Austen[3], Julia Schroeder[1*]

[1] Department of Life Sciences, Imperial College London, Silwood Park, Ascot, SL5 7PY, UK.

[2] School of Pharmacy and Biomolecular Sciences, University of Brighton, Brighton, BN2 4GJ, UK

[3]Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, CT2 7NR UK.

* Corresponding author

The use of crowdsourced data is growing rapidly, particularly in ornithology. Citizen science greatly contributes to our knowledge, however, little is known about the reliability of data collected in that way. We found, using an online picture quiz, that **self-proclaimed expert birders were more likely to misidentify common British bird species as exotic or rare species**, compared to people who rated their own expertise more modestly. This finding suggests that records of rare species should always be considered with caution even if the reporters consider themselves to be experts. In general, however, we show that **self-rated expertise in bird identification skills is a reliable predictor of correct species identification**. Implementing the collection of data on self-rated expertise is easy and low-cost. We therefore suggest it as a useful tool to statistically account for variability in bird identification skills of citizen science participants and to improve the accuracy of identification data collected by citizen science projects.

Introduction

The use of crowdsourced data is growing rapidly (1,2), particularly in ornithology (3). Citizen science data collection (4) greatly contributes to our knowledge of species distribution, population dynamics (4), the assessment of extinction risks (5) and to conservation decision making (6). However, while the correct identification of species is fundamental for the reliability of these data (7) little is known about the variation in the identification skills of the contributors and the so-introduced error. Visual identification is to date still the most efficient and reliable method of most bird species identification (8), yet it relies on the expertise and skill of the observer. Thus, reliance on non-expert species identification, for example in citizen science projects, means that errors will be made. Identification errors can have serious consequences (9). As an example, misidentification of a species that needs to be managed by culling for another one that is endangered (Takahe, *Porphyrio hochstetteri*) can lead to wasted conservation efforts (10). As citizen science data often forms the basis for conservation policies and management plans (6), it is imperative to quantify the extent of these errors. Concerningly, however, few such studies have been conducted. One such rare example is a study showing that expert and non-expert bumblebee species identification are similarly reliable (11), yet experience predicts correct species identification in mussels (7). However, the validity of bird species identification skills remains largely unexplored, and most citizen science projects on birds do not collect information on participants (but see (12)). This is even despite many hobbyist ornithologists contributing to large citizen science projects (13,14). Yet, the popularity of birdwatching (15,16) and the number of people able and willing to contribute to bird citizen science projects bears an immense potential for ornithological research (17). Here, we provide the, to the best of our knowledge, first quantification of visual bird species identification accuracy, with an exceptionally large sample size. We test the hypothesis that people who self-rate their expertise in identifying common bird species higher are also able to correctly identify more birds from pictures. We used an online bird identification questionnaire that presented 2,697 people four pictures of each of six common British bird species.

Results

Descriptive statistics

Our online bird identification questionnaire resulted in 64,728 identification attempts by 2697 potential citizen scientists. We asked participants to rate their own expertise on a five-point scale – self-rated expertise (1 = ‘Novice’, 2 = ‘Little experience with wild birds (feeders in garden, etc.)’, 3 = ‘Intermediate’, 4 = ‘Experience with a wide range of British species, especially common birds’, 5 = ‘Experience with most species in Britain (including waders, gulls, etc.) and abroad (e.g. Western Palearctic)’). We also asked participants whether they had externally certified expertise (e.g. reporting as being trained and licensed as a bird ringer), and of their previous experience in bird surveys. Overall, 78% of the pictured birds were correctly identified.

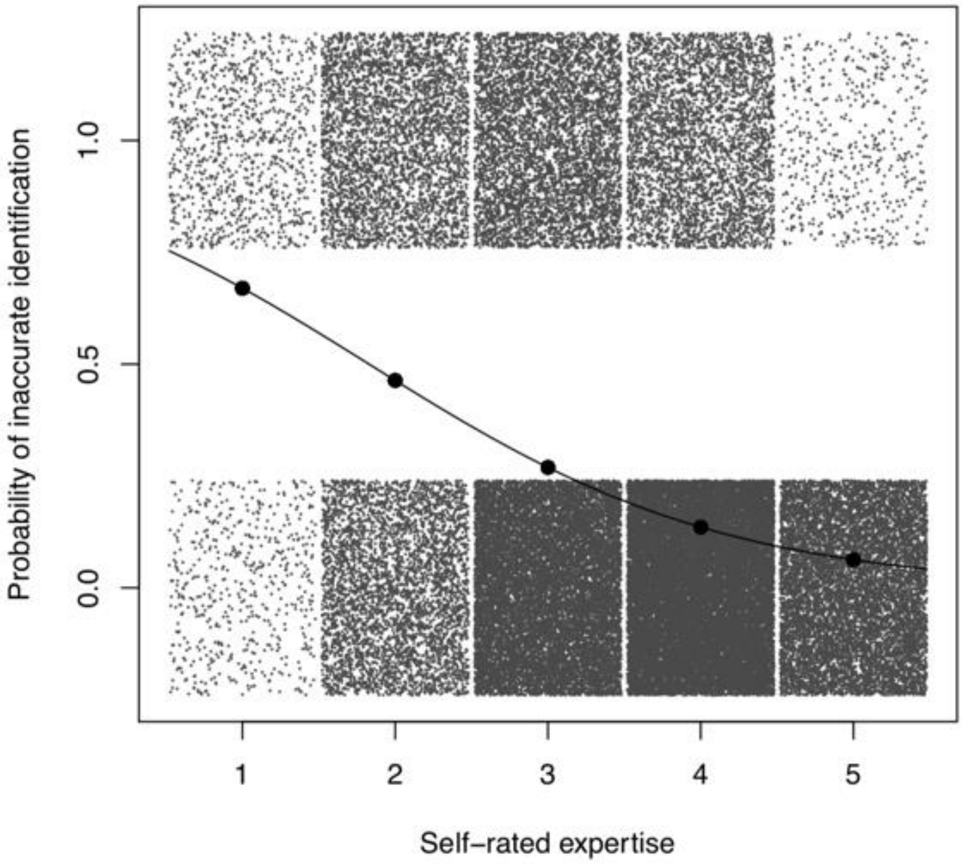

Figure 1: The probability of inaccurate species identification decreases with increasing self-rated expertise, ranging from 1 = Novice to 5 = Expert. The dots represent each one species identification attempt of a single picture (N = 64,728), and are jittered in the x and y directions to visualise sample size per bin. The line and the black filled circles represent predicted values from a Binomial General Linear Model with Identification (0 = correct, 1 = inaccurate) as response variable, and self-rated expertise as explanatory variable.

Self-rated and externally certified expertise as predictors for correct identifications

The probability of an incorrect answer decreased statistically significantly with higher self-rated expertise (Table 1). Self-rated novices (1 on the scale) correctly identified on average 35% of the pictures, while self-rated experts (5 on the scale) correctly identified 95% of all pictures (Fig. 1). While having externally certified expertise and previous experience in bird surveys statistically significantly predicted the probability of correctly identifying a species in a picture, self-rated expertise was a more reliable and precise predictor of correct species identification (Table 1).

Table 1: Higher self-rated bird identification expertise, externally certified expertise, and previous survey expertise all predict fewer inaccurate species identifications. Results from a GLMM of inaccurate species identification (correct = 0, inaccurate = 1) as response variable and self-rated (1=novice, 5=expert), and externally certified (1 = yes, 0 = no), and previous survey experience (1= yes, 0 = no). N = 64,728 species identification attempts of 2,697 participants.

	β	Lower 95CI	Upper 95CI	p
Fixed effects				
Intercept	0.70	0.56	0.83	<0.001
Self-rated expertise	-0.13	-0.14	-0.12	<0.001
Externally certified expertise	-0.03	-0.06	-0.01	0.03
Previous survey experience	-0.04	-0.06	-0.02	<0.001
Random effects	α	Lower 95CI	Upper 95CI	
Participant ID	0.00	0.00	0.00	
Picture ID	0.02	0.01	0.03	
Species	0.02	0.000	0.06	

Incorrect identifications

Inaccurate answers included the acknowledgement of not knowing the answer, and incorrect identifications. Most incorrect identifications referred to other species common in Britain. Surprisingly, despite the title of the questionnaire “Common British birds: identification quiz” and the introductory text explicitly stating that we sought to assess identification skills of common British birds, 113 participants (4.2%) identified at least one of the birds in the pictures as a rarity in Britain, or as a species that has never been reported as wild in Britain (i.e. exotic species, Fig. 2A). Notably, participants who suggested rarities or exotics rated their expertise statistically significantly higher than people who did not suggest rare or exotic bird species, and were also more likely to use references such as bird guide books or websites for help (Fig. 2B). People with higher self-rated expertise are expected to be more familiar with a greater number of species, and therefore may be expected to consider more possible species compared to novices.

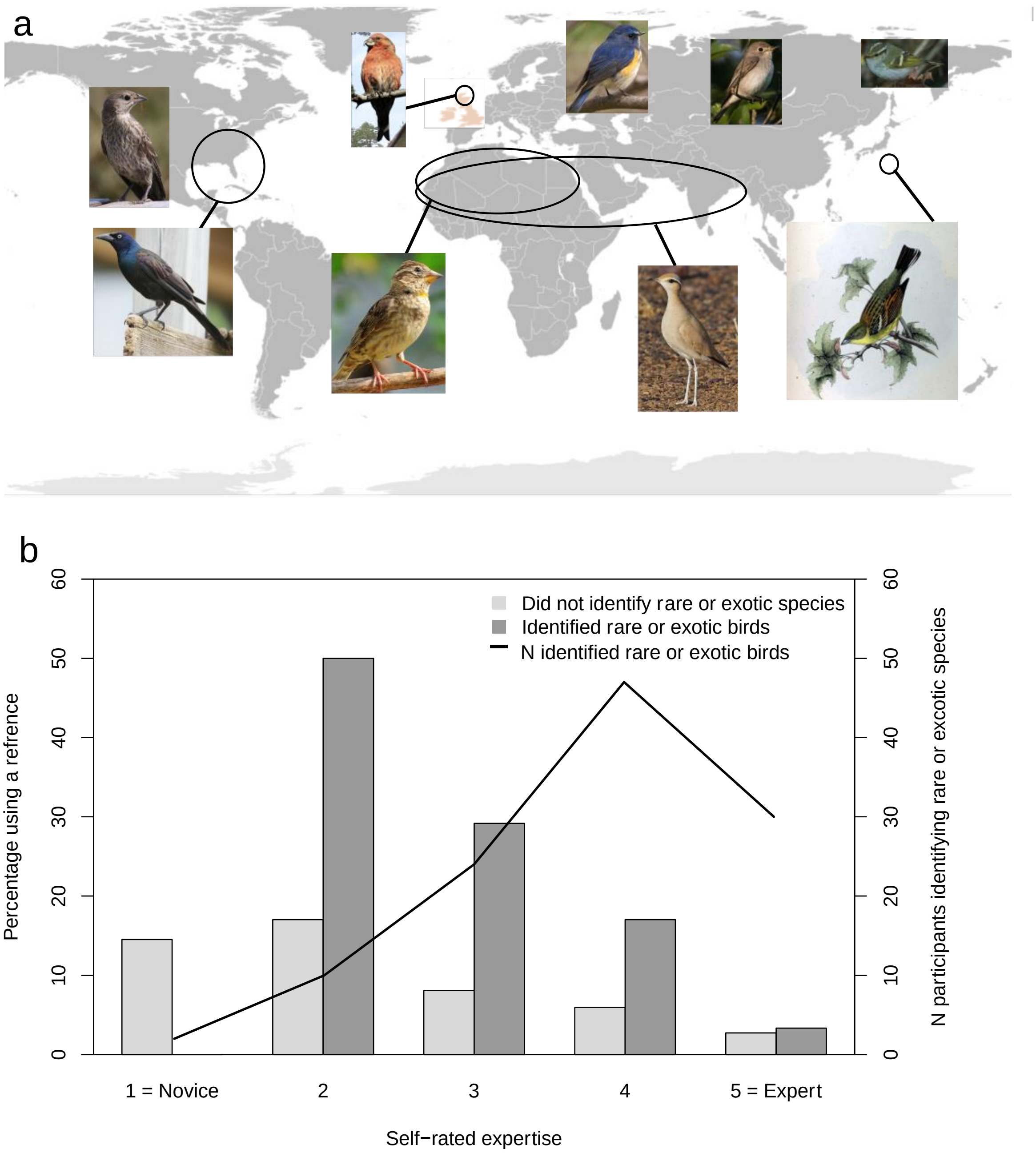

Figure 1 a: A selection of those rare or exotic bird species that participants have most often inaccurately mentioned in the questionnaire. They are placed approximately in the middle of their distribution range, avoiding overlap for visual clarity. The UK map is coloured and enlarged to highlight the crossfinch's range. From left to right and top to bottom: Scottish Crossbill (*Loxia scotica*, photograph by Richard Crossley, cropped. CCA-SA 3.0 license), Redflanked Bluetail (*Tarsiger cyanurus*, photograph by M.Nishimura, cropped, CCA-SA 3.0 license), Pallas's Leaf-warbler (*Phylloscopus proregulus*, photograph by Francesco Veronesi, cropped, CCA-SA 2.0 license), Brown-headed Cowbird (*Molothrus ater*, photograph by Cephas, cropped, CCA-SA 2.0 license), Common Grackle (*Quiscalus quiscula*, by Mdf, CCA-SA 3.0 license), Rock Sparrow (*Petronia petronia*, by Sandra, cropped, CCA-SA 2.0 license), Cream-coloured Courser (*Cursorius cursor*, by Mike Prince, cropped, CCA-SA 2.0 license), Asian Brown Flycatcher (*Muscicapa dauurica*, by Jason Thompson, cropped, CCA-SA 2.0 license) and Yellow Bunting (*Emberiza sulphurata* public domain). Background map: © Sémhur, Wikipedia Commons / CC-BY-SA-3.0.

b: The total number of participants who identified at least one species in a picture as a rare or exotic species (black line, right y-axis). The percentage of participants using reference material like a bird guide book (left y-axis) was higher among participants that inaccurately identified rare or exotic bird species (dark grey bars), than among those that did not identify rare or exotic bird species (light grey bars). Parameter estimates (95CI) of a binomial linear model with rare/exotic species suggested (1 = yes) as response variable: $b_{intercept} = -5.23$ (-6.81– -4.08), $b_{Self\text{-}rated\ Expertise} = 1.34$ (0.78–1.87), $b_{Used\ reference}$ 0.41 (0.20–0.63), N = 2697 participants. Externally certified expertise and previous experience in bird surveys were not associated with seeing rare or exotic species.

Discussion

We found that while in general, self-rated expertise in identifying common bird species did predict the number of correctly identified images, self-rated experts were more likely to identify a common bird species as a rare or exotic species than those people who rated their own expertise more modestly. The incentive of "ticking" (bird watching terminology describing one's first observation of a species) as many species as possible, for a potentially ever growing personal list of observed species, appears to be a common behaviour in birdwatching, although this has not been quantified. There is, to the best of our knowledge, only one study that found no impact of the incentive of personal species list growth on the number of reported false positives, for acoustic bird species identification (18). However, overconfidence certainly could explain the report of a Scottish Crossbill (*Loxia scotica*) in our dataset as this species is not identifiable by sight alone (19). Future research should therefore aim at understanding the underlying causes of the different identification patterns among the different expertise levels.

In conclusion, self-rated expertise is a good indicator of performance and can provide valuable information to any citizen science project involving species identification. We suggest that citizen science projects should evaluate self-rated expertise with a simple questionnaire. The so-collected data can then be used to statistically account for variation in observer expertise, for instance, by using a weighted statistic. We suggest that such an approach should be standard procedure in any citizen science or crowd-sourced project that relies on species identification, to increase precision, reproducibility, and generality of our science.

Materials and methods

Ethics statement

Approval for this study was granted by Prof Barraclough, as representative for the Imperial College Research Ethics Committee. All response forms were anonymous and formal and informed consent was obtained.

Questionnaire

The complete questionnaire is provided as Online Supplementary Information. The selected species were House Sparrow (*Passer domesticus*), Eurasian Blue Tit (*Cyanistes caeruleus*), Common Starling (*Sturnus vulgaris*), European Greenfinch (*Chloris chloris*), Common Chaffinch (*Fringilla coelebs*) and European Robin (*Erithacus rubecula*). No list of possible answers was provided. Pictures for the study species were chosen to reflect natural observation situations in realistic settings, from males, females and juveniles. All used pictures are available in the questionnaire provided in the Online Supplementary Information. The pictures were sourced from the sighting collaborative website observations.be. The plumage differences between British and Belgian birds from the species we selected are negligible (20). We also included one drawing per species that was similar to those presented in bird guide books. The drawings were sourced from the RSPB website with written permission from the artist, Mike

Langman. All participants were informed that the questionnaire only concerned common birds in Great Britain. It was not possible to zoom in on the pictures.

Participant sourcing

Using newsletters ("BTO BirdTrack" and "Wildlife in Ascot"), and social media (Facebook and Twitter), participants were presented a short explanation of the aims of the study and a clarification that all levels of expertise are relevant. The questionnaire was shared on specific Facebook groups targeted to the topic (e.g. UK Bird Identification, Birding UK and Ireland, etc).

Data coding

Species identifications were submitted as free text answers and subsequently checked for spelling mistakes and synonyms and coded using a numeric code (correct, inaccurate). All answers were coded twice and cross-checked to account for human error during coding by NB. Correct species names were accepted even if followed by a question mark, inaccurate sex or similar. Only for the House Sparrow (*Passer domesticus*) was the genus name "sparrow" accepted as a correct answer.

Descriptive statistics

Of all 2697 participants, 66 rated their own expertise as 'Novice' (coded as 1), and 333 described their own expertise as 'Little experience with wild birds (feeders in garden, etc.)' (coded 2). 793 participants considered their own expertise as 'Intermediate' (coded 3), and 1,072 rated themselves as having 'Experience with a wide range of British species, especially common birds' (coded 4). Finally, 433 participants considered themselves experts, described as 'Experience with most species in Britain (including waders, gulls, etc.) and abroad (e.g. Western Palearctic)' (coded 5). We then asked whether participants had previous experience in bird surveys (of which 1,277 (47.3%) participants answered positively) and whether they had been externally certified. We found that 220 participants (7.4%) had either a ringing licence or were a validator on a sighting collection website or similar.

93.3% of all participants were from Britain, 6.1% from other European countries, 0.4% were from outside Europe. Of all participants, 1661 were male, 1018 were female, with 18 participants scored as neither or do not want to say. Only in the self-rated expertise category 4 (‘Experience with a wide range of British species, especially common birds’) was there a significant difference in correctly identifying species in pictures between men and women (two-sided $t = -2.84$, $df = 1068$, $p = 0.005$, all gender comparisons in all other self-rated expertise categories $0.96 > t > -1.68$, and $p > 0.10$). However, note that the data has, due to the large sample size, a high statistical power to discriminate small effect sizes. Here, the effect size was minimal and potentially not biologically important, as women in self-rated expertise category 4 scored on average 20.1 correct out of 24 shown pictures, while men scored 20.7 correctly.

Statistical analysis

To test whether self-rated expertise, externally certified expertise, and previous survey experience predicted the probability of correctly identified bird pictures, we used a generalised linear mixed model (GLMM) with a logit link function. The response variable was either a correctly identified (0) or an inaccurately identified (1) species per picture. The five-level self-rated expertise (1=non-expert, 5=expert) was modelled as a fixed covariate. Externally certified expertise and previous experience were added as two-level fixed factors. Some species may be easier to identify than others. We indeed found that, on average, starlings were least likely to be correctly identified (44% inaccurate identifications), followed by green finch (27%), chaffinch (21%) and house sparrow (18%). Robins (11%) and blue tits to be most likely to be correctly identified (9%). Therefore, we modelled species as a random effect. To account for variation between participants and to account for pseudo-replication, we modelled participant ID as a random effect on the intercept. We accounted for the fact that some pictures may have been easier to identify than others by modelling picture ID as a random effect on the intercept. We found a statistically significant difference between the probability to correctly identify a drawing and a photograph (χ^2-test: $\chi^2 = 114.8$, df = 1, $p < 0.0001$). Note that the low p-value stems from the large sample size and thus high statistical power to detect small effects. Indeed, the actual difference between both categories was minimal (% inaccurately identified: photos 21.9%, drawings 21.0%) and likely irrelevant. However, the random effect of picture ID statistically corrects for any difference between photos and drawings. We used Bayesian Mixed Models and R package MCMCglmm (21) to model GLMMs, these account well for over-dispersion in the data. We used an inverse Wishart prior for the random effects. The residual variance is not identifiable when using binary data, therefore, we used the prior to fix it to 1. The models were run with 75,000 iterations and the default burn-in parameter. We report posterior means as parameter estimates, and 95% credible intervals. We used a t-test to test whether people who reported rare or non-British birds had higher self-rated expertise. All analyses were conducted in R version 3.5.0 (22).

Supplementary Information. The complete questionnaire can be found here: https://docs.google.com/forms/d/e/1FAIpQLSeBIqWcy4YPBGf6YeDwApVCR0od6FBXSoVpDKYpsN5fmz9tIg/viewform

Ethics

Approval for this study was granted by DHoD Prof Barraclough, as representative for the Imperial College Research Ethics Committee. All response forms were anonymous and formal and informed consent was obtained.

References

1. Cohn JP. Citizen science: Can volunteers do real research? BioScience. 2008 Mar;58(3):192–7.
2. Williams RL, Stafford R, Goodenough AE. Biodiversity in urban gardens: Assessing the accuracy of citizen science data on garden hedgehogs. Urban Ecosystems. Springer US; 2015 Sep;18(3):819–33.
3. Dickinson JL, Zuckerberg B, Bonter DN. Citizen Science as an Ecological Research Tool: Challenges and Benefits. Annu Rev Ecol Evol Syst. Annual Reviews; 2010;41(1):149–72.
4. Harris SJ. The breeding bird survey 2016. bto.org. 2017.
5. Solow A, Smith W, Burgman M, Rout T, Wintle B, Roberts D. Uncertain Sightings and the Extinction of the Ivory-Billed Woodpecker. Cons Biol. Wiley/Blackwell (10.1111); 2012 Feb;26(1):180–4.
6. Sutherland WJ, Roy DB, Amano T. An agenda for the future of biological recording for ecological monitoring and citizen science. Biol J Linn Soc. 2015 Jul;115(3):779–84.
7. Shea CP, Peterson JT, Wisniewski JM, Johnson NA. Misidentification of freshwater mussel species (Bivalvia:Unionidae): contributing factors, management implications, and potential solutions. Journal of the North American Benthological Society. The University of Chicago Press; 2011 Jun;30(2):446–58.
8. Handley LL. How will the "molecular revolution' contribute to biological recording? Biol J Linn Soc. 2015 Jul;115(3):750–66.
9. Dennhardt AJ, Duerr AE, Brandes D, Katzner TE. Integrating citizen-science data with movement models to estimate the size of a migratory golden eagle population. Biol Conserv. 2015 Apr;184:68–78.

10. Department of Conservation (DOC), DOC asks police to consider suspending licences after takahē shooting (2015). http://www.doc.govt.nz/news/media-releases/2015/doc-asks-police-to-consider-suspending-licences-after-takahe-shooting/>. 2015.

11. Austen GE, Bindemann M, Griffiths RA, Roberts DL. Species identification by experts and non-experts: comparing images from field guides. Sci Rep. 2016;6(1).

12. Comber A, Mooney P, Purves RS, Rocchini D, Walz A. Crowdsourcing: It Matters Who the Crowd Are. The Impacts of between Group Variations in Recording Land Cover. Matisziw TC, editor. PLoS one. 2016;11(7).

13. Reynolds MD, Sullivan BL, Hallstein E, Matsumoto S, Kelling S, Merrifield M, et al. Dynamic conservation for migratory species. Science Advances. American Association for the Advancement of Science; 2017 Aug;3(8).

14. Sullivan BL, Aycrigg JL, Barry JH, Bonney RE, Bruns N, Cooper CB, et al. The eBird enterprise: An integrated approach to development and application of citizen science. Biol Conserv. 2014 Jan;169:31–40.

15. Sali MJ, Kuehn DM. Exploring motivations among male and female non-residential birdwatchers in New York State. Human Dimensions of Wildlife. Taylor & Francis Group; 2008 May 1;13(3):201–2.

16. Rothery L, Scott GW, Morrell LJ. Colour preferences of UK garden birds at supplementary seed feeders. Dyer AG, editor. PLoS one. 2017;12(2).

17. Kelling S, Lagoze C, Wong W-K, Yu J, Damoulas T, Gerbracht J, et al. eBird: A Human/Computer Learning Network to Improve Biodiversity Conservation and Research. Ai Magazine. 2013;34(1):10–20.

18. Farmer RG, Leonard ML, Horn AG. Observer Effects and Avian-Call-Count Survey Quality: Rare-Species Biases and Overconfidence. Auk. University of California Press; 2012 Jan;129(1):76–86.

19. Birds AKB, 1990. Identification of crossbill and Scottish crossbill. britishbirdscouk.

20. Svensson L, Mullarney K, Zetterström D, Grant PJ. Collins Bird Guide. Collins; 2011. 1 p.

21. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. J Stat Softw. 2010;33(2):1–22.

22. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria.