

1 **IDENTIFYING CHARISMATIC BIRD SPECIES AND TRAITS**  
2 **WITH COMMUNITY SCIENCE OBSERVATIONS**

3 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

4 <sup>1</sup> Statistical and Data Sciences Program, Smith College

5 <sup>2</sup> Department of Environmental Science, Policy, and Management,  
6 University of California, Berkeley

7 \* sstoudt@berkeley.edu

8 † equal contribution

9 **This is a pre-print.**

10 **ABSTRACT**

11 Identifying which species are perceived as charismatic can improve the impact  
12 and efficiency of conservation outreach, as charismatic species receive more conser-  
13 vation funding and have their conservation needs prioritized (9; 17; 13). Sociologi-  
14 cal experiments studying animal charisma have relied on stated preferences to find  
15 correlations between hypothetical “willingness to pay” or “empathy” for a species’  
16 conservation and species’ size, color, and aesthetic appeal (51; 13; 16). Recognizing  
17 the increasing availability of digital records of public engagement with animals that  
18 reveal preferences, an emerging field of “culturomics” uses Google search results,  
19 Wikipedia article activities, and other digital modes of engagement to identify  
20 charismatic species and traits (46; 31; 10; 41). In this study, we take advantage  
21 of community science efforts as another form of digital data that can reveal ob-  
22 server preferences. We apply a multi-stage analysis to ask whether opportunistic  
23 birders contributing to iNaturalist engage more with larger, more colorful, and

2 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

24 rarer birds relative to a baseline, from eBird contributors, approximating unbiased  
25 detection. We find that body mass, color contrast, and range size all predict over-  
26 representation in the opportunistic dataset. We also find evidence that, across 473  
27 modeled species, 52 species are significantly overreported and 158 are significantly  
28 underreported, indicating a wide variety of species-specific effects. Understanding  
29 which birds are charismatic can aid conservationists in creating impactful outreach  
30 materials and engaging new naturalists. The quantified differences between two  
31 prominent community science efforts may also be of use for researchers leveraging  
32 the data from one or both of them to answer scientific questions of interest.

### 33 1. INTRODUCTION

34 Birds have received special attention in conservation (21; 48), and investigations  
35 into stated preferences for birds found that species traits—color, pattern, and  
36 shape—influence their perceived charisma (9; 37; 38). Others have taken advantage  
37 of revealed-preference data for birds from the volume of Google search results  
38 and the Common Breeding Bird Monitoring Scheme (62) and eBird data (50) to  
39 similarly investigate the public’s perception of different birds.

40 Online community science platforms, which collect data contributed by volun-  
41 teers, provide a more direct way to study public perception of species in the wild.  
42 Community scientists, sometimes called “citizen scientists,” volunteer contribu-  
43 tions to scientific databases as self-guided, non-professionals. Two biodiversity  
44 platforms in particular are of great interest for investigating bird charisma on a  
45 large scale. eBird, an app for hobbyist birders that has generated one of the  
46 world’s largest biodiversity databases, has recorded over 550 million records in  
47 North America to date (22). Many of these records come from over 46 million  
48 “complete checklists” (20) and thus represent a rigorous reporting protocol, with

49 reliable information on when species were not seen alongside when they were seen  
50 as well as the inclusion of sampling metadata (18). Another popular platform  
51 is iNaturalist, a nature app designed to encourage public engagement with all  
52 species. The primary goal of iNaturalist is to “connect people to nature” (28).  
53 The app allows any observation of any species at any time or place to be entered,  
54 so reporting rates depend on relative interest in different species.

55 Biodiversity records such as those aggregated on online community science plat-  
56 forms are important for informing species distribution models (54; 14; 43; 56).  
57 Complete checklists from eBird are lauded as appropriate for species distribu-  
58 tion modeling (11; 53), whereas opportunistic records such as from iNaturalist are  
59 known to contain particular biases (29; 40). These biases are often characterized as  
60 noise, but they may actually contain a strong signal of the habits and preferences  
61 of opportunistic naturalists. In this paper, we estimate those biases in relation  
62 to eBird and thereby analyze public interest in species traits such as size, color  
63 contrast, phylogenetic relatedness, and taxonomic order. By harnessing two huge  
64 but very different community science datasets, we gain insight into human interest  
65 in biodiversity as encountered in the wild.

66 We construct a conceptual model to relate eBird and iNaturalist’s data-generating  
67 processes and show how they can be studied to characterize observer biases and  
68 preferences (Figure 1). Imagine an iNaturalist user who sees a bird, takes a pic-  
69 ture of it, and submits the photo to iNaturalist. For this event to occur, three  
70 separate conditions must be satisfied. First, the species must be present in the  
71 environment. We call this condition the “presence” filter, and characterizing this  
72 process is the main goal of most species distribution models that use community  
73 science data. Second, the observer must see the species—this is the “detectability”  
74 filter, which is controlled for in ecological studies as imperfect detection. Were this

4

SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

75 eBird, the process would stop here, because all detected birds must be reported  
76 under the complete-checklist protocol. However, in iNaturalist, a third event must  
77 occur: having seen the species, the observer must then make the choice to record  
78 an audio clip or, far more commonly, take a photo and upload it to iNaturalist. We  
79 call this the “interest” filter. Characterizing the interest filter, visible as deviations  
80 in iNaturalist reporting rates relative to eBird, is the main goal of this study.

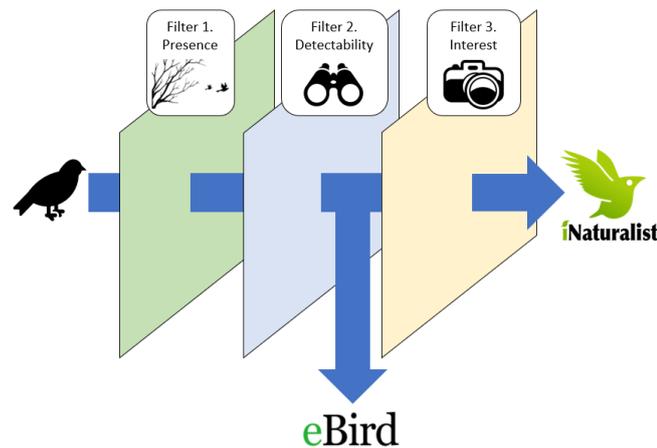


FIGURE 1. Community science reporting as a filtering process. eBird data passes through two conceptual filters: (1) presence and (2) detectability. In iNaturalist, a hypothetical report must pass through an additional filter, (3) interest.

81

## 2. MATERIALS AND METHODS

82 2.1. **eBird and iNaturalist pre-processing and data structures.** In August  
83 2020 we downloaded the eBird Basic Dataset and extracted all complete checklists.  
84 We also obtained all iNaturalist research grade observations of bird species from  
85 the Global Biodiversity Information Facility (23). For both datasets, we considered  
86 only observations made in the contiguous 48 United States and Washington, D.C.  
87 on or before December 31, 2019. We excluded checklists obtained in 2020 or  
88 later to avoid accidentally capturing changes in observer behavior related to the

89 COVID-19 pandemic (25; 15). We associated species across the two datasets using  
90 the R package `taxalight` (8).

91 We aggregated observations to a spatial grid of regular hexagonal cells covering  
92 the contiguous U.S. such that spatial grid cells had a long radius (center to vertex)  
93 of 20 kilometers. In each “hex,” a count was obtained for the number of times each  
94 species was detected. The total number of observations in each hex was calculated  
95 as the sum of the species observation events in that hex. For consistency with  
96 iNaturalist, an eBird checklist that reported more than one individual of a species  
97 was counted as a series of separate species observation events (a checklist with  
98 three species reported constituted three observations). By aggregating over time  
99 we assumed that the primary variation is spatial. We accounted for secondary  
100 variation elsewhere in the modeling approach.

101 We sub-selected species according to the following criteria. First, to capture  
102 spatial variability in sampling, we only considered species that were observed one  
103 or more times in at least 100 different hexes. Second, we only considered species in  
104 the EltonTraits database (60; 58). We eliminated “pelagic specialists” according  
105 to EltonTraits, expecting the sampling process generating these data to be funda-  
106 mentally different from that of terrestrial birds. 482 North American bird species  
107 met these criteria.

108 **2.2. Species-level spatial analysis.** We first estimated a typical overreporting  
109 index characterizing each species. The unit of analysis for the first stage was  
110  $y_{ijk}$ , the count of reports of species  $j$  reported at the  $i$ th spatial hex in dataset  $k$   
111 (either eBird or iNaturalist). We modeled the number of “successes” in a binomial  
112 random draw  $y_{ijk}$  as  $y_{ijk} \sim \text{Binom}(C_{ik}, R_{ijk})$ . Here  $C_{ik}$  is the known number of  
113 “attempts,” which we define for the presence-only iNaturalist data to be the total

114 number of iNaturalist observations across all bird species in spatial hex  $i$  and for  
115 the eBird checklist data to be the total number of species observation events on all  
116 of the eBird checklists in spatial hex  $i$ . (Note that for eBird, a single checklist on  
117 which 5 different species are observed would be considered five species observation  
118 events, rather than one, in order to match the iNaturalist sampling schema.)  $R_{ijk}$   
119 is the reporting rate which we model as a function of location  $i$  for each species  $j$   
120 and dataset  $k$  combination.

121 We used a quasi-binomial generalized additive model (GAM) with a logit link  
122 to capture spatial variability via a multidimensional tensor-product smooth of the  
123 longitude and latitude coordinates of the hexes (61). The motivation for this  
124 approach was twofold. First, we anticipated that many differences between the  
125 datasets could be due to spatial heterogeneity in sampling. Both eBird and iNat-  
126 uralist are highly spatially variable with their own hotspots—for example, eBird  
127 is most densely used near its base in Cornell, NY, and on the East Coast, while  
128 iNaturalist has high density regions in California and Texas (19; 27). For this  
129 reason alone we expect that reporting rates would vary by dataset, and so a spa-  
130 tially explicit analysis is called for. A spatially explicit approach is also necessary  
131 since user habits may themselves be spatially non-independent. To obtain accurate  
132 confidence intervals on parameter estimates, spatial autocorrelation in the data-  
133 generating process must be accounted for. Second, we anticipated extra-binomial,  
134 non-spatial variability across units (e.g. temporal, weather conditions, observer  
135 variability, etc.). We chose a quasi-binomial approach as a way to account for this  
136 in the uncertainty quantification.

137 A GAM was fit for each species and dataset combination. The basis dimensions  
138 were chosen to be 20 knots by 20 knots. We fit all models and then iteratively  
139 increased each dimension by 5 until the model passed a hypothesis test of whether

140 the basis dimension for a smooth was adequate using a p-value cutoff  $\alpha = 0.1$   
141 (given in the R function `mgcv::gam.check`).

142 From the quasi-binomial GAMs, we obtained estimates of the spatially smooth  
143 surface of the reporting rate at each hex, in each dataset. We calculated the  
144 overreporting index as the median predicted difference in log-scale reporting rates  
145 across hexes for each species. To obtain accurate confidence intervals on the over-  
146 reporting index, we used a parametric bootstrap approach, making random draws  
147 of the spatial surface and recomputing the index each time, to obtain an estimate  
148 of uncertainty for each index. Each species' overreporting index represents the  
149 typical deviation in the iNaturalist reporting rate relative to the eBird baseline for  
150 that species.

151 To assess which overreporting indices were significantly different from zero, we  
152 used a p-value threshold that was adjusted to account for the fact that we made  
153 multiple comparisons (one for each species). We used a false discovery rate con-  
154 trolling method to ensure that across comparisons the false positive rate was no  
155 more than 0.05 (7).

156 **2.3. Cross-species meta-analysis.** To answer our main question, whether species  
157 traits can help explain these differences in birder engagement, we used a meta-  
158 analysis to explain patterns in the overreporting indices. A meta-analysis allowed  
159 us to propagate the uncertainty estimated in the first stage of the analysis (57).  
160 The median differences in reporting rates for all of the species, along with their  
161 standard errors, became the response in this stage of the analysis.

162 We hypothesized two types of charisma-associated bias that could drive differ-  
163 ences between the datasets. The first, photogenic bias, refers to sampling dif-  
164 ferences due to aesthetic preferences of observers. One component of photogenic

165 bias relates to relevant aspects of species charisma, such as size and color. To  
166 investigate the effect of size, we retrieved species' log mass from the EltonTraits  
167 dataset (58). To represent how colorful or striking a bird is, we used an index of  
168 maximum color contrast originally developed by Schuetz and Johnston (50). An-  
169 other component of photogenic bias is logistical: particular bird species might be  
170 overreported by virtue of being easier to document rather than more charismatic.  
171 Since iNaturalist observations are almost always associated with photos this could  
172 lead to a difference between the datasets. The second type of bias, novelty bias,  
173 may occur where users in iNaturalist preferentially report species that are new to  
174 them or that they see infrequently. We used two covariates as proxies of different  
175 aspects of rarity: the number of hexes a species is reported in (a proxy for size of  
176 effective range) and the proportion of all eBird checklists where the species was  
177 found (a spatially agnostic proxy for overall prevalence). We centered and scaled  
178 these covariates for log mass, maximum color contrast, log range size, and log  
179 prevalence.

180 In the meta-analysis we also incorporated phylogenetic structure to account for  
181 the possibility that phylogenetically closer species have more similar reporting in-  
182 dices due to evolutionary non-independence of unmodeled but important traits  
183 (1; 36). We obtained multiple phylogenetic trees from BirdTree.org (30). We then  
184 obtained a consensus tree including branch edges using the R package phytools  
185 (47). Finally we computed a variance-covariance matrix based on this consen-  
186 sus tree using the R package ape (45). We allowed for both a random effect for  
187 species with this variance-covariance structure and an unstructured random ef-  
188 fect for species, similar to an approach of creating a mixture of a phylogenetic  
189 covariance matrix and the identity matrix (44).

190 We fit a second meta-analysis including only the effect of taxonomic order and  
191 excluding phylogenetic structure to obtain estimates of each order's mean overre-  
192 porting index with properly propagated error (4; 33).

193 Models for nine species failed to fit (see Appendix) and were therefore dropped  
194 from the second stage analysis. Three of these failed to pass a test for adequate  
195 knots with a basis dimension of 35 by 35, above which computation became in-  
196 feasible. Six of these failed to converge in under 24 hours, which was chosen as a  
197 practical cutoff.

198 We removed 49 species that had overreporting indices outside the range -10 to  
199 10 from the meta-analysis stage. Values less than -10 or greater than 10 arose  
200 in cases where, among the union of hexes where a species was reported in either  
201 dataset, one dataset reported no observations in over half of those hexes. Because  
202 the reporting index uses median differences in log reporting rates, it could not  
203 be reliably estimated, nor its uncertainty reliably quantified for the meta-analysis  
204 step, in these cases. After these two filters, 424 species were included. To test  
205 the sensitivity of results, we replicated the meta-analysis including the extreme  
206 overreporting indices.

207

### 3. RESULTS

208 We predicted an overreporting index with uncertainty for each of 473 species of  
209 interest. Even when controlling the false discovery rate, 210 species had overreport-  
210 ing indices significantly different from zero, giving evidence that iNaturalist and  
211 eBird reporting rates were meaningfully different for many species. A significantly  
212 negative overreporting index means iNaturalist observers were uninterested, while  
213 a significantly positive one means they were interested. Figure 2 shows the most  
214 extreme over- and under-reported birds. The most overreported birds are each

10 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

215 some combination of large (wild turkey), well-known for their appeal (burrowing  
216 owl) or considered especially beautiful (Indian peafowl). The least overreported  
217 birds follow less of an obvious pattern, though they tend to be smaller or associated  
218 with near-ocean foraging.

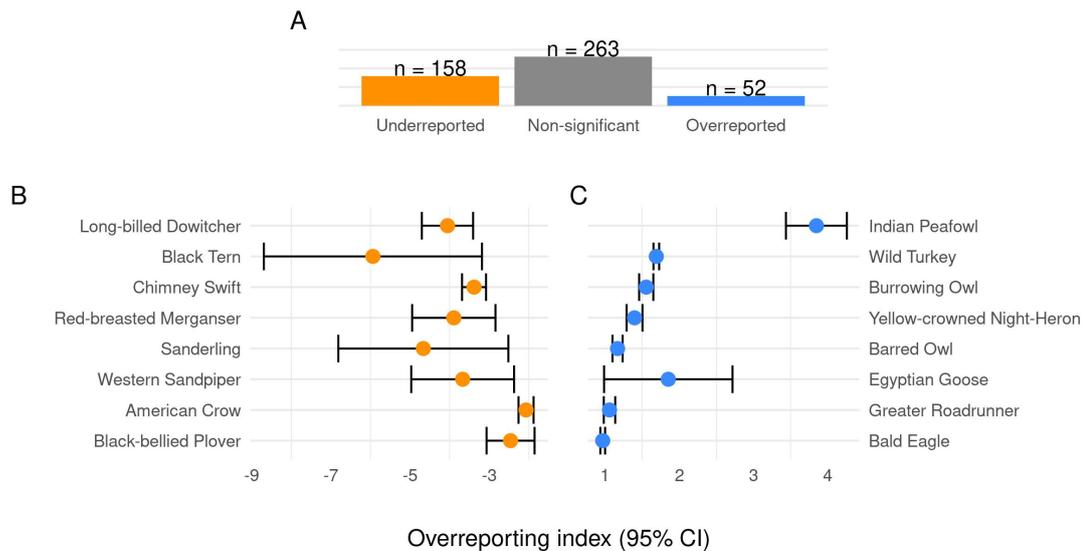


FIGURE 2. Species overreporting indices. (A) Counts of bird overreporting indices by significance level, controlling the FDR to account for multiple comparisons. The most under-reported species (B) appeared in iNaturalist between 0.002-0.13 as frequently as in eBird, while the most over-reported species (C) appeared in iNaturalist roughly 1.6-45 times as frequently as in eBird. Over-reporting index is median difference in log reporting rate. For each species, the estimate with 95% CI is shown.

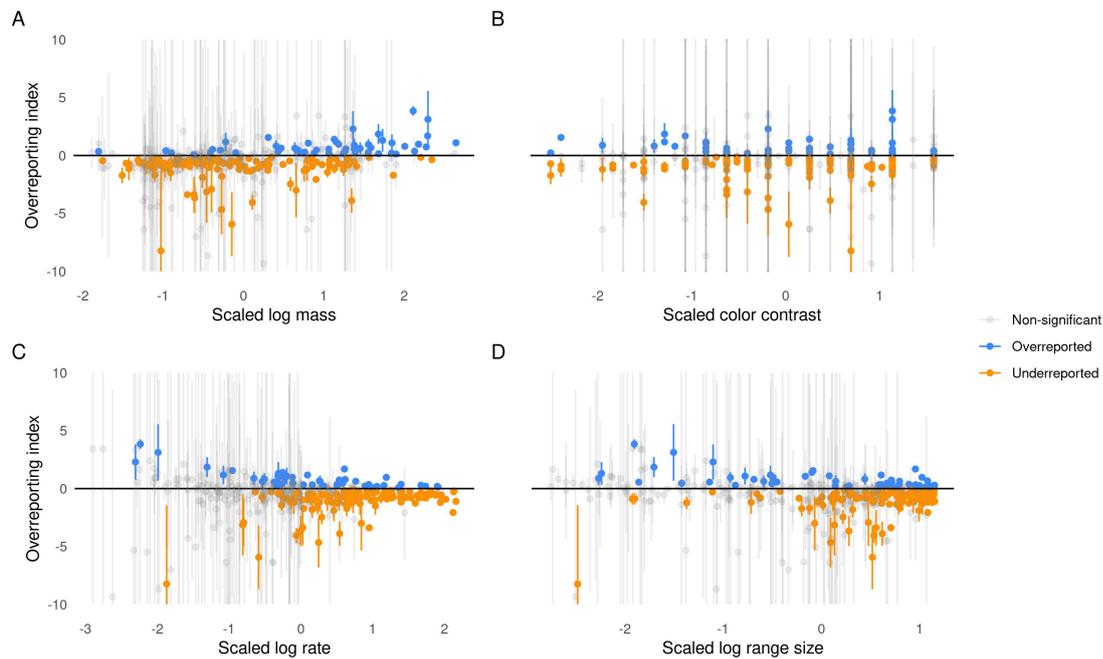


FIGURE 3. Species overreporting indices plotted against four traits—(A) log mass, (B) color contrast, (C) log prevalence, and (D) log range size—whose associations were studied in a meta-analysis. Vertical error bars show 95% confidence intervals of indices, which do not correspond exactly to the significance test adjusted for multiple testing.

219 We hypothesized that these differences at the species level, represented by the  
220 third filter layer (Figure 1), could be driven by a variety of mechanisms. We  
221 next investigated whether photogenic bias and novelty bias were present. Figure 3  
222 shows the relationship between overreporting indices and these species traits that  
223 relate to the proposed biases.

224 We found a statistically significant relationship between the overreporting index  
225 and size, color, and number of hexes containing observations of the species (a proxy  
226 for range size) in a second stage meta-analysis. This means that opportunistic  
227 birders in the contiguous U.S. chose larger, more colorful, and more range-limited  
228 birds more often than would be expected based on the corresponding species'

12 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

229 detection rates in eBird. We found a 0.31 effect size on scaled log mass with a  
230 95% confidence interval of (0.15, 0.47), 0.12 for scaled color (0.04, 0.19), -0.13 for  
231 scaled log number of hexes where reported (-0.25, -0.03), and -0.08 for scaled log  
232 proportion of eBird checklists where reported (-0.19, 0.03). See Table 1 in the  
233 Appendix for more interpretation. The independent random effect dominates the  
234 phylogenetically structured random effect in magnitude, suggesting there is not  
235 much phylogenetic structure in the residual variance.

236 We tested sensitivity of results to the choice to exclude low information indices  
237 with extreme values. Including these indices made the effect of overall prevalence  
238 significantly positive, removed the significance of color, and increased the coeffi-  
239 cients for mass and the overall prevalence (See Appendix for details). Thus, results  
240 are sensitive to an attempt to include unreliable indices for species with limited  
241 data. See the Appendix for further discussion of these points.

242 We accounted for phylogeny at the species level in the meta-analysis, but it  
243 can also be helpful to visualize relationships at the order level for improved inter-  
244 pretation. We identified five taxonomic orders whose typical overreporting indices  
245 were significantly different from zero after correcting for multiple testing (Figure 4).  
246 Owls and gamefowl tend to be relatively large, which may explain their overreport-  
247 ing. However, many waders and gulls are large, and this order is underreported,  
248 along with songbirds and doves.

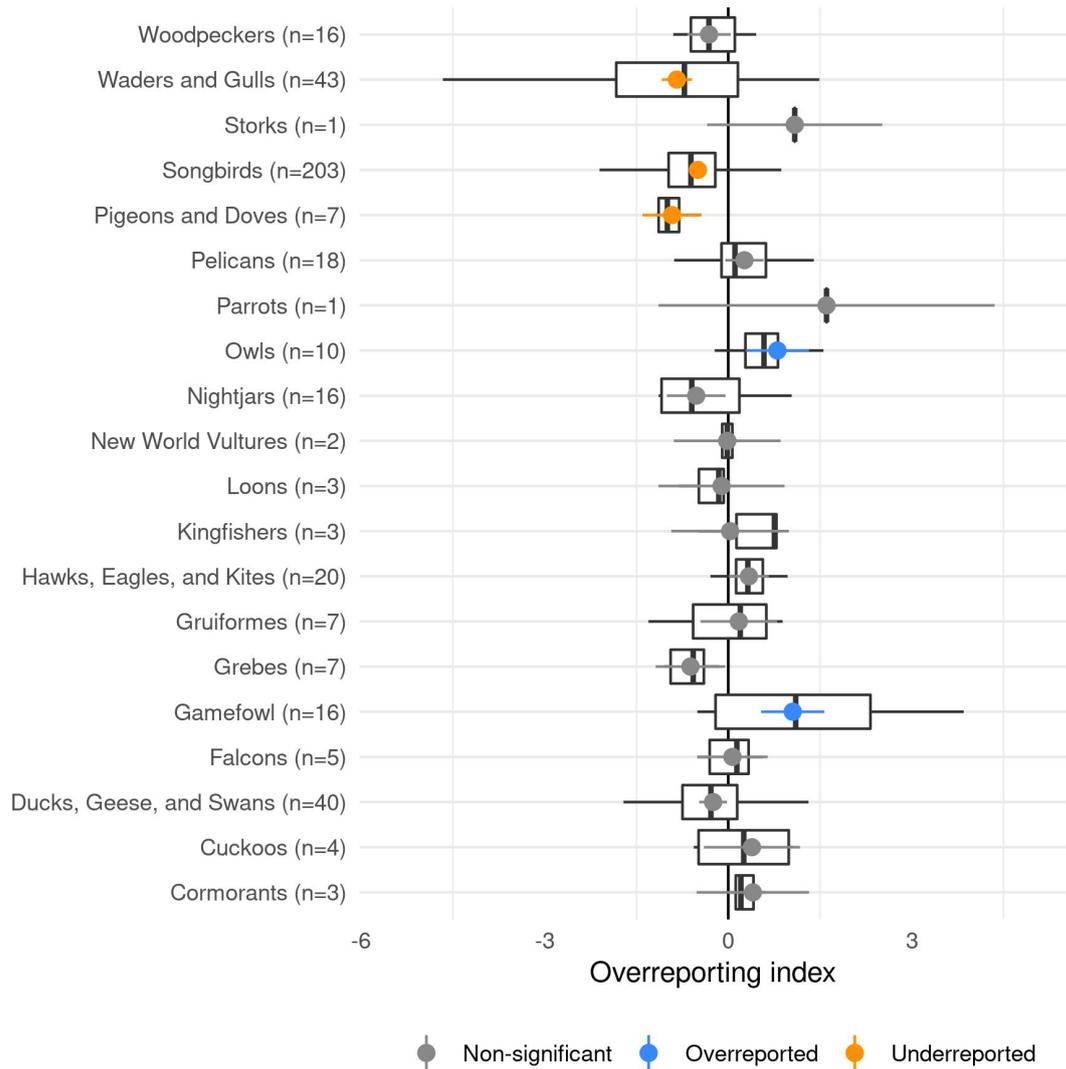


FIGURE 4. The predicted effect of order, representing the typical overreporting index for that group of species, excluding low-information outliers. Colors indicate the results of a false discovery rate-controlling significance test. Box plots show the range of estimated overreporting indices in each group.

249

#### 4. DISCUSSION

250 When using community science data to study biodiversity, the inescapable fact of  
251 variation between users is usually treated as noise that at best adds uncertainty and

252 at worst causes biased estimation. However, when studying people's relationships  
253 to animals rather than the animals themselves, with an eye towards using this  
254 knowledge to inform conservation efforts (35), the paradigm flips, and variation  
255 between users becomes the data-generating process of interest. For this task,  
256 community science is not merely a source of noisy data, but a unique and invaluable  
257 source of information about how members of the public choose to engage with the  
258 natural world.

259 The iNaturalist dataset in particular is ideal for characterizing naturalists' pref-  
260 erences. Because iNaturalist is open to all living organisms, anything like eBird's  
261 complete checklist protocol is impossible, as an observer will likely see hundreds of  
262 plants and dozens of animals for every individual they choose to report. iNatural-  
263 ist therefore forces its users to constantly discern which organisms they consider  
264 noteworthy and which can be ignored. This process reveals the human element of  
265 choice.

266 To obtain estimates of birder interest for 473 bird species, we modeled variation  
267 in iNaturalist reports across species relative to the eBird baseline. To increase  
268 confidence that any findings reflect true differences in reporting behaviors, we  
269 included four critical layers of statistical rigour intended to estimate unbiased  
270 reporting rates and accurately estimate uncertainty: a spatially explicit model,  
271 a quasi-binomial error distribution, a parametric bootstrap on the overreporting  
272 index, and a phylogenetically explicit meta-analysis.

273 While preparing this manuscript for submission, we became aware of indepen-  
274 dent work conducted simultaneously by Callaghan et al. (12) who answer a similar  
275 question using different statistical methods and an alternative overreporting esti-  
276 mate.

277 In the wild, opportunistic birders engage more with species whose traits have  
278 been identified as charismatic in artificial contexts such as surveys (13; 16). In  
279 particular, this study reinforces the concept of the charismatic megafauna, the idea  
280 that larger species are more interesting or sympathetic to the public (32; 17; 34).  
281 The finding that more range-restricted birds are overrepresented is also consistent  
282 with the hypothesis that iNaturalist users may optimize for lengthening their “life  
283 list”, the list of unique species they have ever observed on the app(39). This  
284 “gotta catch ’em all” commodification of nature can be at odds with the scientific  
285 pursuits that this data informs (2). The strong relationship we find between size  
286 and overreporting is also consistent with the hypothesis that iNaturalist users  
287 are drawn to birds that are easier to photograph. Logistical constraints around  
288 photography are likely not the only drivers of variation, as hinted by the fact that  
289 the American crow, a common, large, and relatively bold bird, is quite easy to  
290 photograph but is strongly underreported.

291 Nature photography and film-making have played a large role in conservation  
292 by bringing biodiversity to the attention of the public (5; 52; 24). As camera  
293 lenses and photos have become ubiquitous in our culture, conservation sites and  
294 museums have leveraged this fact, using Instagram and other social media plat-  
295 forms to understand and further engage visitors, both by producing aesthetically  
296 compelling imagery and by encouraging visitors to take and share their own pho-  
297 tographs (59; 55; 26). iNaturalist takes advantage of commonplace camera phones  
298 to provide users with automatic identifications of their observations (49) and a  
299 platform on which to document and share their experiences. The logistical and  
300 charisma aspects of this interest bias, seen through a camera lens, are likely corre-  
301 lated, and we were not able to disentangle them within this framework. However,  
302 because photographs and imagery play a large role in modern communication and

16 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

303 social media, logistical obstacles to photographing a bird may play a large role in  
304 how that bird is known or perceived by community members and therefore pat-  
305 terns in logistical bias may themselves be of interest. Quantifying the difference  
306 between these two community science platforms may also be of interest to scien-  
307 tists using the data from these sources, and future work could include leveraging  
308 these differences in modeling approaches.

309 It is important to note that both eBird and iNaturalist are not uniformly ac-  
310 cessible across user demographics, so these results can only reveal the preferences  
311 of people who are already engaging with nature through these community science  
312 platforms, and who have self-selected based on a general interest in birds. Future  
313 work could investigate the “interest” filter in further sub-population analyses (e.g.  
314 children, urban areas) (6; 3) or with respect to seasonality (42).

315 Understanding which birds are the most charismatic can aid conservationists  
316 in creating impactful outreach materials, whether it be due to their size, color,  
317 rarity, or myriad other cultural and species-specific factors. Conservationists may  
318 tailor outreach materials, arguments, and experiences to best align with people’s  
319 preferences. This knowledge is also useful for engaging new naturalists who may  
320 not be represented in the initial analysis; by knowing which birds people tend  
321 to most enjoy seeing, an organizer can tailor a birding trip to include the most  
322 charismatic species.

323 ACKNOWLEDGEMENTS

324 Thanks to S. Beissinger, C. Boettiger, M. Chapman, J. Clare, and W. Fithian  
325 for comments and support, and to the dedicated users of eBird and iNaturalist who  
326 provided data for this study. SS was supported by the Gordon and Betty Moore  
327 Foundation through Grant GBMF3834, and by the Alfred P. Sloan Foundation

328 through Grant 2013-10-27 to the University of California, Berkeley. BG was sup-  
329 ported by the National Science Foundation Graduate Research Fellowship under  
330 Grant No. 1752814. This work used the Extreme Science and Engineering Discov-  
331 ery Environment (XSEDE), which is supported by National Science Foundation  
332 Grant No. ACI-1548562.

333 REFERENCES

- 334 [1] D. C Adams. Phylogenetic meta-analysis. *Evolution*, 62(3):567–572, 2008.  
335 <https://doi.org/10.1111/j.1558-5646.2007.00314.x>.
- 336 [2] S. Altrudi. Connecting to nature through tech? The case of the iNat-  
337 uralist app. *Convergence: The International Journal of Fresearch into*  
338 *New Media Technologies*, 27(1):124–141, 2021. [https://doi.org/10.1177/](https://doi.org/10.1177/1354856520933064)  
339 [1354856520933064](https://doi.org/10.1177/1354856520933064).
- 340 [3] M. Aristeidou, C. Herodotou, H. L. Ballard, A. N. Young, A. E. Miller, L. Hig-  
341 gins, and R. F. Johnson. Exploring the participation of young citizen scien-  
342 tists in scientific research: The case of iNaturalist. *PLoS ONE*, 16(1), 2021.  
343 <https://doi.org/10.1371/journal.pone.0245682>.
- 344 [4] G. Arnqvist and D. Wooster. Meta-analysis: synthesizing research findings in  
345 ecology and evolution. *Trends in Ecology & Evolution*, 10(6):236–240, 1995.  
346 [https://doi.org/10.1016/S0169-5347\(00\)89073-4](https://doi.org/10.1016/S0169-5347(00)89073-4).
- 347 [5] T. A. Barbas, S. Paraskevopoulos, and A. G. Stamou. The effect of na-  
348 ture documentaires on environmental sensitivity: a case study. *Learn-*  
349 *ing, Media, and Technology*, 34(1):61–69, 2009. [https://doi.org/10.1080/](https://doi.org/10.1080/17439880902759943)  
350 [17439880902759943](https://doi.org/10.1080/17439880902759943).
- 351 [6] J. A. Belaire, L. M. Westphal, C. J. Whelan, and E. S. Minor. Urban residents’  
352 perceptions of birds in the neighborhood: Biodiversity, cultural ecosystem

- 353 services, and disservices. *The Condor*, 117(2):192–202, 2015. <https://doi.org/10.1650/CONDOR-14-128.1>.
- 354
- 355 [7] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57(1):289–300, 1995. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- 356
- 357
- 358
- 359 [8] C. Boettiger and K Norman. *taxalight: A Lightweight and Lightning-Fast Taxonomic Naming Interface*, 2021. R package version 0.1.3.
- 360
- 361 [9] M. Brambilla, M. Gustin, and C. Celada. Species appeal predicts conservation status. *Biological Conservation*, 160:209–213, 2013. <https://doi.org/10.1016/j.biocon.2013.02.006>.
- 362
- 363
- 364 [10] Z. Burivalova, R. A. Butler, and D. S. Wilcove. Analyzing Google search data to debunk myths about the public’s interest in conservation. *Frontiers in Ecology and the Environment*, 16(9):509–514, 2018. <https://doi.org/10.1002/fee.1962>.
- 365
- 366
- 367
- 368 [11] C. T. Callaghan and D. E. Gawlik. Efficacy of ebird data as an aid in conservation planning and monitoring. *Journal of Field Ornithology*, 86(4):298–304, 2015.
- 369
- 370
- 371 [12] C. T. Callaghan, A. G. B. Poore, M. Hofmann, C. Roberts, and H. M. Pereira. Large-bodied birds are over-represented in opportunistic citizen science data. *bioRxiv*, 2021. <https://ecoevorxiv.org/vnspb/>.
- 372
- 373
- 374 [13] A. Colleony, S. Clayton, D. Couvet, M. Saint Jalme, and A. Prevot. Human preferences for species conservation: Animal charisma trumps endangered status. *Biological Conservation*, 206:263–269, 2016. <https://doi.org/10.1016/j.biocon.2016.11.035>.
- 375
- 376
- 377

- 378 [14] C. L. Coxen, J. K. Frey, S. A. Carleton, and D. P. Collins. Species distribution  
379 models for a migratory bird based on citizen science and satellite tracking data.  
380 *Global Ecology and Conservation*, 11:298–311, 2017. [https://doi.org/10.](https://doi.org/10.1016/j.gecco.2017.08.001)  
381 [1016/j.gecco.2017.08.001](https://doi.org/10.1016/j.gecco.2017.08.001).
- 382 [15] T. M. Crimmins, E. Posthumus, S. Schaffer, and K. L Prudic. COVID-19  
383 impacts on participation in large scale biodiversity-themed community science  
384 projects in the United States. *Biological Conservation*, 256:109017, 2021.  
385 <https://doi.org/10.1016/j.biocon.2021.109017>.
- 386 [16] P. Curtin and S. Papworth. Coloring and size influence preferences for imagi-  
387 nary animals, and can predict actual donations to species-specific conservation  
388 charities. *Conservation Letters*, 13(4):e12723, 2020. [https://doi.org/10.](https://doi.org/10.1111/conl.12723)  
389 [1111/conl.12723](https://doi.org/10.1111/conl.12723).
- 390 [17] F. Ducarme, G. M. Luque, and F. Courchamp. What are “charismatic species”  
391 for conservation biologists? *BioSciences Master Reviews*, 2013.
- 392 [18] eBird. About eBird. <https://ebird.org/about>.
- 393 [19] eBird. Hotspot. <https://ebird.org/hotspots>.
- 394 [20] eBird. North America. <https://ebird.org/region/na>.
- 395 [21] S. T. Garnett, G. B. Ainsworth, and K. K. Zander. Are we choosing the right  
396 flagships? the bird species and traits australians find most attractive. *PLoS*  
397 *ONE*, 13(6), 2018. <https://doi.org/10.1371/journal.pone.0199253>.
- 398 [22] Global Biodiversity Information Facility (GBIF). EOD -  
399 eBird observation dataset. [https://www.gbif.org/dataset/](https://www.gbif.org/dataset/4fa7b334-ce0d-4e88-aaae-2e0c138d049e)  
400 [4fa7b334-ce0d-4e88-aaae-2e0c138d049e](https://www.gbif.org/dataset/4fa7b334-ce0d-4e88-aaae-2e0c138d049e).
- 401 [23] Global Biodiversity Information Facility (GBIF). GBIF occurrence download.  
402 <https://doi.org/10.15468/dl.7mt5mn>.

20 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

- 403 [24] E. Hanisch, R. Johnston, and N. Longnecker. Cameras for conservation:  
404 wildlife photography and emotiaonal engagement with biodiversity and na-  
405 ture. *Human Dimensions of Wildlife*, 24(3):267–284, 2019. [https://doi.](https://doi.org/10.1080/10871209.2019.1600206)  
406 [org/10.1080/10871209.2019.1600206](https://doi.org/10.1080/10871209.2019.1600206).
- 407 [25] W. M. Hochachka, H. Alonso, C. Gutierrez-Exposito, E. Miller, and A. John-  
408 ston. Regional variation in the impacts of the COVID-19 pandemic on the  
409 quantity and quality of data collected by the project eBird. *Biological Con-*  
410 *servation*, 254:108974, 2021. [https://doi.org/10.1016/j.biocon.2021.](https://doi.org/10.1016/j.biocon.2021.108974)  
411 [108974](https://doi.org/10.1016/j.biocon.2021.108974).
- 412 [26] K. Hughes and G. Moscardo. Connecting with new audiences: Exploring the  
413 impact of mobile communication devices on the experiences of young adults  
414 in museums. *Visitor Studies*, 20(1):33–55, 2017. [https://doi.org/10.1080/](https://doi.org/10.1080/10645578.2017.1297128)  
415 [10645578.2017.1297128](https://doi.org/10.1080/10645578.2017.1297128).
- 416 [27] iNaturalist. Observations. <https://www.inaturalist.org/observations>.
- 417 [28] iNaturalist. What is it. <https://www.inaturalist.org/pages/what+is+it>.
- 418 [29] N. J. B. Isaac and M. J. O. Pocock. Bias and information in biological records.  
419 *Biological Journal of the Linnean Society*, 115(3):522–531, 2015. [https://](https://doi.org/10.1111/bij.12532)  
420 [doi.org/10.1111/bij.12532](https://doi.org/10.1111/bij.12532).
- 421 [30] W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. A global  
422 phylogeny of birds. <https://birdtree.org/subsets/>.
- 423 [31] J. Y. Kim, Y. Do, R. Im, G. Kim, and G. Joo. Use of large web-based data  
424 to identify public interest and trends related to endangered species. *Biodi-*  
425 *versity and Conservation*, 23:2961–2984, 2014. [https://doi.org/10.1007/](https://doi.org/10.1007/s10531-014-0757-8)  
426 [s10531-014-0757-8](https://doi.org/10.1007/s10531-014-0757-8).

- 427 [32] E. Kneegtering, H. J. Van der Windt, and A. J. M. Schoot Uiterkamp. Public  
428 decisions on animal species: does body size matter? *Environmental Conser-*  
429 *vation*, 38(1):28–36, 2010. <https://www.jstor.org/stable/44519558>.
- 430 [33] J. Koricheva, J. Gurevitch, and K. Mengersen, editors. *Handbook of Meta-*  
431 *analysis in Ecology and Evolution*. Princeton University Press, Princeton,  
432 New Jersey, 2013.
- 433 [34] M. Krause. Charismatic species and beyond: how cultural schemas and organ-  
434 isational routines shape conservation. *Conservation and Society*, 15(3):313–  
435 321, 2017. <https://www.jstor.org/stable/26393299>.
- 436 [35] R. J. Ladle, R. A. Correia, Y. Do, G. Joo, A. C. M. Malhado, R. Proulx,  
437 J. Roberge, and P. Jepson. Conservation culturomics. *Frontiers in Ecology and*  
438 *the Environment*, 14:269–275, 2016. <https://doi.org/10.1002/fee.1260>.
- 439 [36] M. J. Lajeunesse. Meta-analysis and the comparative phylogenetic method.  
440 *The American Naturalist*, 174(3):369–381, 2009. [https://doi.org/10.1086/](https://doi.org/10.1086/603628)  
441 603628.
- 442 [37] S. Liskova and D. Frynta. What determines bird beauty in human  
443 eyes? *Anthrozoos*, 26(1):27–41, 2013. [https://doi.org/10.2752/](https://doi.org/10.2752/175303713X13534238631399)  
444 175303713X13534238631399.
- 445 [38] S. Liskova, E. Landova, and D. Frynta. Human preferences for colorful birds:  
446 Vivid colors or pattern? *Evolutionary Psychology*, 13(2):147470491501300203,  
447 2015.
- 448 [39] S. Loarie. A new kind of life list. [https://www.inaturalist.org/blog/](https://www.inaturalist.org/blog/42454-a-new-kind-of-life-list)  
449 42454-a-new-kind-of-life-list.
- 450 [40] C. Meyer, P. Weigelt, and H. Kreft. Multidimensional biases, gaps and uncer-  
451 tainties in global plant occurrence information. *Ecology Letters*, 19:992–1006,  
452 2016. <https://doi.org/10.1111/ele.12624>.

- 453 [41] J. C. Mittermeier, R. Correia, R. Grenyer, T. Toivonen, and U. Roll. Using  
454 Wikipedia to measure public interest in biodiversity and conservation. *Con-*  
455 *servation Biology*, 35(2):412–423, 2021. [https://doi.org/10.1111/cobi.](https://doi.org/10.1111/cobi.13702)  
456 13702.
- 457 [42] J. C. Mittermeier, U. Roll, T. J. Matthews, and R. Grenyer. A season for all  
458 things: Phenological inprints in Wikipedia usage and their relevance to con-  
459 servation. *PLoS Biology*, 17(3), 2019. [https://doi.org/10.1371/journal.](https://doi.org/10.1371/journal.pbio.3000146)  
460 [pbio.3000146](https://doi.org/10.1371/journal.pbio.3000146).
- 461 [43] K. Pacifici, B. J. Reich, D. A. W. Miller, B. Gardner, G. Stauffer, S. Singh,  
462 A. McKerrow, and J. A. Collazo. Integrating multiple data sources in species  
463 distribution modeling: a framework for data fusion. *Ecology*, 98(3):840–850,  
464 2017. <https://doi.org/10.1002/ecy.1710>.
- 465 [44] M. Pagel. Detecting correlated evolution on phylogenies: a general method  
466 for the comparative analysis of discrete characters. *Proceedings of the Royal*  
467 *Society of London. Series B: Biological Sciences*, 255(1342):37–45, 1994.  
468 <https://doi.org/10.1098/rspb.1994.0006>.
- 469 [45] E. Paradis and K. Schliep. ape 5.0: an environment for modern phylogenetics  
470 and evolutionary analyses in R. *Bioinformatics*, 35:526–528, 2019.
- 471 [46] R. Proulx, P. Massicotte, and M. Pepino. Googling trends in conservation  
472 biology. *Conservation Biology*, 28(1):44–51, 2013. [https://doi.org/10.](https://doi.org/10.1111/cobi.12131)  
473 [1111/cobi.12131](https://doi.org/10.1111/cobi.12131).
- 474 [47] L. J. Revell. phytools: an R package for phylogenetic comparative biology  
475 (and other things). *Methods in Ecology and Evolution*, 3:217–223, 2012.
- 476 [48] S. K. Robinson. Bird niches in human culture and why they matter. *Pro-*  
477 *ceedings of the National Academy of Sciences of the United States of America*,  
478 116(22):10620–10622, 2019. <https://doi.org/10.1073/pnas.1905901116>.

- 479 [49] C. Saari. iNaturalist computer vision explorations. [https://www.inaturalist.org/pages/computer\\_vision\\_demo](https://www.inaturalist.org/pages/computer_vision_demo).
- 480
- 481 [50] J. G. Schuetz and A. Johnston. Characterizing the cultural niches of North  
482 American birds. *Proceedings of the National Academy of Sciences of the United  
483 States of America*, 116(22):10868–10873, 2019. [https://doi.org/10.1073/  
484 pnas.1820670116](https://doi.org/10.1073/pnas.1820670116).
- 485 [51] J. A. Serpell. Factors influencing human attitudes to animals and their welfare.  
486 *Animal Welfare*, 13:S145–S152, 2004.
- 487 [52] M. J. Silk, S. L. Crowley, A. J. Woodhead, and A. Nuno. Considering connec-  
488 tions between hollywood and biodiversity conservation. *Conservation Biology*,  
489 32(3):597–606, 2017. <https://doi.org/10.1111/cobi.13030>.
- 490 [53] V. A. Steen, C. S. Elphick, and M. W. Tingley. An evaluation of stringent  
491 filtering to improve species distribution models from citizen science data. *Di-  
492 versity and Distributions*, 25:1857–1869, 2019. [https://doi.org/10.1111/  
493 ddi.12985](https://doi.org/10.1111/ddi.12985).
- 494 [54] S. R. Supp, F. A. La Sorte, T. A. Cormier, M. C. W. Lim, D. R. Powers, S. M.  
495 Wethington, S. Goetz, and C. H. Graham. Citizen-science data provides new  
496 insight into annual and seasonal variation in migration patterns. *Ecosphere*,  
497 6(1):1–19, 2015. <https://doi.org/10.1890/ES14-00290.1>.
- 498 [55] H. Tenkanen, E. Di Minin, V. Heikinheimo, A. Hausmann, M. Herbst, L. Ka-  
499 jala, and T. Toivonen. Instagram, Flickr, or Twitter: Assessing the usability of  
500 social media data for visitor monitoring in protected areas. *Scientific Reports*,  
501 7(1):1–11, 2017. <https://doi.org/10.1038/s41598-017-18007-4>.
- 502 [56] C. Van Eupen, D. Maes, M. Herremans, K. R. R. Swinnen, B. Somers, and  
503 S. Luca. The impact of data quality filtering of opportunistic citizen sci-  
504 ence data on species distribuiton model performance. *Ecological Modelling*,

24 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

505 444:109453, 2021. <https://doi.org/10.1016/j.ecolmodel.2021.109453>.

506 [57] W. Viechtbauer. Conducting meta-analyses in R with the metafor package.

507 *Journal of Statistical Software*, 36(3):1–48, 2010. [10.18637/jss.v036.i03](https://doi.org/10.18637/jss.v036.i03).

508 [58] J. Walter. Dataset: Elton traits. [http://www.esapubs.org/archive/ecol/](http://www.esapubs.org/archive/ecol/E095/178/)  
509 [E095/178/](http://www.esapubs.org/archive/ecol/E095/178/).

510 [59] A. Weilenmann, T. Hillman, and B. Jungselius. Instagram at the museum:  
511 Communicating the museum experience through social photo sharing. In  
512 *Proceedings of the SIGCHI conference on human factors in computing systems*,  
513 pages 1843–1852, 2013. <https://doi.org/10.1145/2470654.2466243>.

514 [60] H. Wilman, J. Belmaker, J. Simpson, C. de la Rosa, M. M. Rivadeneira,  
515 and W. Jetz. EltonTraits 1.0: Species-level foraging attributes of the world’s  
516 birds and mammals: Ecological archives E095-178. *Ecology*, 95(7):2027, 2014.  
517 <https://doi.org/10.1890/13-1917.1>.

518 [61] S. N. Wood. Fast stable restricted maximum likelihood and marginal like-  
519 lihood estimation of semiparametric generalized linear models. *Journal of*  
520 *the Royal Statistical Society Series B*, 73:3–36, 2011. [https://doi.org/10.](https://doi.org/10.1111/j.1467-9868.2010.00749.x)  
521 [1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).

522 [62] M. Zmihorski, J. Dziarska-Palac, T. H. Sparks, and P. Tryjanowski. Eco-  
523 logical correlates of the popularity of birds and butterflies in Internet infor-  
524 mation resources. *Oikos*, 122:183–190, 2013. [https://doi.org/10.1111/j.](https://doi.org/10.1111/j.1600-0706.2012.20486.x)  
525 [1600-0706.2012.20486.x](https://doi.org/10.1111/j.1600-0706.2012.20486.x).

526

APPENDIX

527 The nine species dropped due to convergence issues in the first stage were:  
528 chestnut-sided Warbler, double-crested cormorant, Swainson's thrush, red cross-  
529 bill, black-throated blue warbler, western kingbird, Nashville warbler, ash-throated  
530 flycatcher, and Cassin's kingbird.

531 Figure 1 helps motivate the exclusion of the small number of species that had  
532 extremely negative overreporting indices. Without these values, the distribution  
533 of overreporting indices is less dramatically left-skewed.

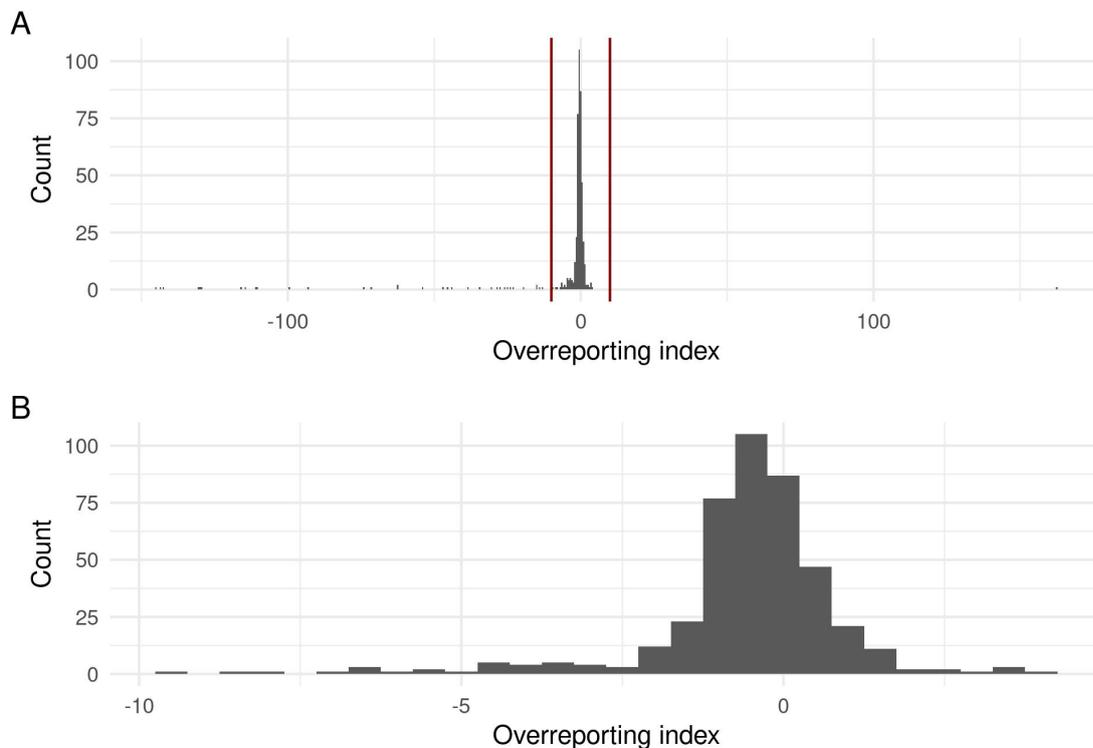


FIGURE 1. Histograms of overreporting indices before and after filtering show that the bulk of the estimates are distributed near 0, with extreme values being spread over a large range. Most, but not all, of these extremely negative values were associated with large uncertainties. Red lines indicate the cutoff  $(-10, 10)$ . See the Materials and methods section in the main manuscript for a discussion of how to interpret these extremely negative values.

534 Table 1 aims to aid the interpretation of results from the cross-species meta-  
535 analysis when low-information outliers were excluded. Mass, color, and the proxy  
536 for species range were all significant. The effects for mass and color are positive  
537 while the effect for the species range proxy is negative.

<b>Covariate</b>	<b>Effect Size</b>	<b>Interpretation</b>
Scaled log mass	0.31 (0.15, 0.47)	An increase by one standard deviation of bird size on the log scale is associated with an expected 36% increase in the odds of overreporting.
Scaled color	0.12 (0.04, 0.19)	An increase by one standard deviation of bird color contrast is associated with an expected 12% increase in the odds of overreporting.
Scaled log number of hexes where reported	-0.13 (-0.25, -0.03)	An increase by one standard deviation of number of hexes on the log scale is associated with an expected 11% decrease in the odds of overreporting.
Scaled log proportion of eBird checklists where found	-0.08 (-0.19, 0.03)	An increase by one standard deviation of proportion of eBird checklists where a species is found, on the log scale, is associated with a 8% decrease in the odds of overreporting.

TABLE 1. Meta-analysis coefficients and interpretations for species traits once low information outliers were excluded (overreporting indices greater than 10 or less than -10). All covariates except for the proxy of overall prevalence were found to be statistically significant.

538 The meta-analysis results are impacted by the inclusion of the 49 low information  
539 outliers. Table 2 aims to aid the interpretation of results from the meta-analysis  
540 when low information outliers were included. The significance of color is lost and

541 the significance of the proxy for overall prevalence is gained. The signs for mass  
 542 and proxy for spatial range do not change though they get much more extreme.

<b>Covariate</b>	<b>Effect Size</b>	<b>Interpretation</b>
Scaled log mass	4.81 (0.49, 9.14)	An increase by one standard deviation of bird size on the log scale is associated with an expected nearly 23-fold increase in overreporting.
Scaled color	0.23 (-1.47, 1.92)	An increase by one standard deviation of bird color contrast is associated with an expected 26% decrease in the odds of overreporting.
Scaled log number of hexes where spotted	-4.35 (-6.50, -2.19)	An increase by one standard deviation of number of hexes on the log scale is associated with an expected nearly 78-fold decrease in the odds of overreporting.
Scaled log proportion of eBird checklists where found	5.96 (3.61, 8.31)	An increase by one standard deviation of proportion of eBird checklists where a species is found, on the log scale, is associated with a nearly 388-fold increase in overreporting.

TABLE 2. Meta-analysis coefficients and interpretations for species traits. All but color are statistically significant, but the effect sizes are very extreme on this scale, alluding to instability.

543 The 49 low-information outliers were: Abert's towhee, alder flycatcher, Arc-  
 544 tic tern, Bachman's sparrow, bank swallow, Bicknell's thrush, black rosy-finch,  
 545 black swift, black-chinned sparrow, boreal owl, Cassin's vireo, Connecticut war-  
 546 bler, Cordilleran flycatcher, dusky flycatcher, Eastern whip-poor-will, elf owl,  
 547 flammulated owl, gray flycatcher, greater sage grouse, Hammond's flycatcher,  
 548 Henslow's sparrow, hepatic tanager, hoary redpoll, Hudsonian godwit, Kentucky  
 549 warbler, king eider, king rail, Lucy's warbler, mountain plover, mountain quail,

28 SARA STOUDT<sup>1\*†</sup>, BENJAMIN R. GOLDSTEIN<sup>2†</sup>, PERRY DE VALPINE<sup>2</sup>

550 Nelson's sparrow, parasitic jaeger, Philadelphia vireo, pinyon jay, red knot, red  
551 phalarope, red-cockaded woodpecker, sage thrasher, Smith's longspur, Stejneger's  
552 scoter, stilt sandpiper, Swainson's warbler, Vaux's swift, Virginia's warbler, white-  
553 headed woodpecker, Williamson's sapsucker, yellow rail, yellow-bellied flycatcher,  
554 and yellow-billed loon.

555 These low-information outliers are generally extreme overreporting indices. These  
556 primarily arise from cases of few or no iNaturalist observations in many of the hexes  
557 where the bird was reported in eBird. This could be partially explained by the fact  
558 that iNaturalist has far fewer total bird observations such that many rare or un-  
559 derreported species will never appear in the dataset in certain hexes, and therefore  
560 a rate difference cannot be reliably estimated in spite of the fact that a parametric  
561 bootstrap of median differences consistently predicts an extreme difference.