PNAS
www.pnas.org

Supplementary Information for

Hoping for Optimality or Designing for Inclusion: Persistence, Learning and the Social Network of Citizen Science

Julia K. Parrish[1], Timothy Jones[1], Hillary K. Burgess[1], Yurong He[1], Lucy Fortson[2] and Darlene Cavalier[3]

[1] School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195-5020

[2] School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455

[3] School for the Future of Innovation in Society, Arizona State University, Tempe, AZ 85287-5603

Corresponding author: Julia K. Parrish

Email: jparrish@uw.edu

This file includes:

Supplementary text

Figs. S1 to S5

Tables S1 to S2

References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1807186116

Supplementary Text

COASST demographics

Complete demographic information is known for 54% (N=1,270) of regular COASST participants, including:

1. Gender - male, female
2. Age at training - categorized for this analysis into 3 levels: <40, 40-60 and >60
3. Self-reported bird identification expertise - no experience, beginner, intermediate, advanced, and expert; categorized for this analysis as one of three levels: 1: no experience/beginner, 2: intermediate, or 3: expert/advanced
4. Membership in other conservation, environmental or citizen science groups; categorized for this analysis into a binomial factor: yes, no.

COASST retention

Retention curves were generated from duration information on participant involvement (extracted on 3 May 2018) generated from first and last survey date. We restricted the participant set by excluding all currently active participants (defined as completing a survey within the last year), as calculation of participant-specific retention was only possible for individuals with a firm end point.

We generated project-wide retention curves out to 10 years, bootstrapping 95% confidence intervals by sampling 70% of the population (1,000 permutations). Retention curves were also calculated as a function of demographic and participant experience categories (gender, age class, birding expertise, previous citizen science involvement, sociality).

For long-term participants (5+ years, and including currently active participants; N = 491), we examined the relationship between effort (defined as total number of surveys, and carcasses found, respectively) and persistence by fitting a generalized linear model (GLM; negative binomial, identity link function) modeling number of surveys or carcasses as a linear function of persistence. Correlations were determined with pseudo-R^2 values as a measure of explanatory power:

$$R^2 = 1 - \frac{Deviance_{Residual}}{Deviance_{Null}} \quad \text{(eq. S1)}$$

COASST data collection accuracy

For this analysis, accuracy was defined as the percent of carcasses identified correctly to species plus the percent identified correctly as "species unknown." The latter category is operationally defined as carcasses for which a species-level identification is impossible (according to the verifier) but a larger taxonomic grouping is possible (e.g., murres, large grebes, true puffins). All accuracy analyses are based on carcass finds from 2013 onwards, and exclude unusual mortality events (surveys with >30 carcasses), when carcass abundance dictated an alternate, abbreviated protocol.

We examined accuracy at three different levels: cumulative participant accuracy, carcass-specific accuracy and survey-specific accuracy. *Cumulative participant accuracy* is participant-specific accuracy over all birds encountered by that individual, regardless of participant team size, and was assessed for N = 774 participants who began surveying after 1-Jan-2013.

Carcass-specific accuracy was used to generate population-level learning curves by fitting a negative exponential curve describing the probability of successful identification to species (binomial: yes/no) as a function of previous COASST experience, proxied as cumulative carcasses found by the most experienced member of the relevant team. The probability of correctly identifying a bird, A, was modelled as a function of previous bird finds, b', according to:

$$A(b') = A_0 + d(1 - e^{-\gamma b'}) \qquad \text{(eq. S2)}$$

where A_0 models initial accuracy, d is the difference between initial and asymptotic accuracy and γ models learning rate. All birds were binned according to previous bird finds, *b'* expressed in integer increments (i.e. b' = 0, 1, 2, 3 …) and eq. S2 was fitted to the binned data assuming that carcass-specific accuracy varied according to a binomial distribution, with probability of correct identification equal to *A(b')*, with n_{trials} = bin-specific number of carcasses, and $n_{successes}$ = number of carcasses correctly identified in that bin.

Learning curve models were estimated using STAN in R (1) assuming non-informative/flat priors for all three parameters (A_0, d, γ) of the negative exponential model. Parameter estimates were obtained from 4 MCMC chains, with a burn-in period of 5,000 iterations, or until model convergence as determined by visual examination of cross-chain mixing of parameter estimates. Upon convergence, chains were run for a further 5,000 iterations to estimate model parameters.

We also fitted learning curves to each of the demographic factor categories using the modeling approach above. Carcasses were assigned to demographic categories (e.g., all participants >60 years of age at training), according to the demographics expressed by the most experienced member of the survey team, judged based on previous number of carcass finds. When team members had equivalent experience, carcass finds were only assigned demographic information if those participants were also in the same demographic class, and otherwise excluded from the demographic specific analyses.

At the *survey-level* we examined the influence that three factors: carcass abundance, number of species found, and previous experience (proxied by the participant-specific maximum of cumulative carcass finds) on within-survey learning. As for carcass-specific accuracy, each survey was assigned to the most experienced participant on that survey. We fit generalized additive models (GAM) using the mgcv package in R (2), assuming that the number of carcasses correctly identified in a survey, $n_{correct}$, varied according to a binomial distribution with n_{trials} equal to the survey-specific bird abundance, and survey-specific probability of success informed by smooth functions of survey-specific bird abundance and diversity as well as previous beached bird experience. GAMs consisting of all possible combinations of these factors were fitted and ranked according to AIC, with the best model selected as the one that minimized AIC. Predictor importance was determined by examination of model ranks and by comparing AIC values of reduced models (i.e. best model, with one factor excluded) to the overall best model. A bootstrap resampling routine (1,000 models, each constructed on 70% of the full dataset selected at random) was used to calculate the mean and 95% CI of the fitted relationships resulting from the best model, and to visualize the relationship between accuracy and each of the model predictors.

Accuracy-retention trade-offs

We examined trade-offs in retention and accuracy in two ways. At the project scale we generated a trade-off frontier, or the line describing estimated retention at a given level of accuracy. Conversions between these values were made by selecting a region with a non-zero carcass encounter rate, allowing previous experience (i.e. carcasses found) at a given accuracy level to be converted to months, assuming 12 surveys annually. For this example we use the Pacific Northwest outer coast where on average 32 carcasses are typically found annually (assuming one survey per calendar month) with 50% of beaches recording between 14 and 43 carcasses. Representative of this region we assumed that each individual survey represented 3 new carcass finds, with 30 days between surveys.

To examine trade-offs specific to particular demographics, we extracted short-term (one survey) and long-term (one year) measures of retention and accuracy from the demographic-specific accuracy and retention curves (i.e., Fig. S4) as follows: short-term retention = participants completing 2+ surveys (i.e. percent repeaters); long-term retention = participants persisting for 1+ year; short-term accuracy = fitted carcass-specific accuracy for the first carcass encountered; long-term accuracy = fitted carcass-specific accuracy at 30 birds. The latter threshold was selected to be representative of expected participant accuracy after one year of surveying Pacific Northwest outer coast beaches, where typically ~ 30 carcasses are encountered per year (see above).

Uncertainty was calculated as the model-fitted 95% confidence interval of the predicted mean accuracy at 1 and 30 birds, respectively, and uncertainty in retention at these two points in time calculated via resampling (1,000 permutations each sampling 70% of the participants belonging to that demographic class).

Social and network analysis

We created an adjacency matrix, with matrix values $M_{i,j}$ corresponding to the number of surveys in common between individuals i and j, and used this to visualize the COASST social network using the R package igraph (3). We used these networks to understand the distribution of network sizes, participant abundance at size, and the relative occurrence of nexus people. Finally, we used the networks to estimate the number of new individuals recruited by existing participants by attributing all novice (first survey) partners to all more experienced survey team members (and where multiple assignations were possible for team sizes >2) as a "recruit."

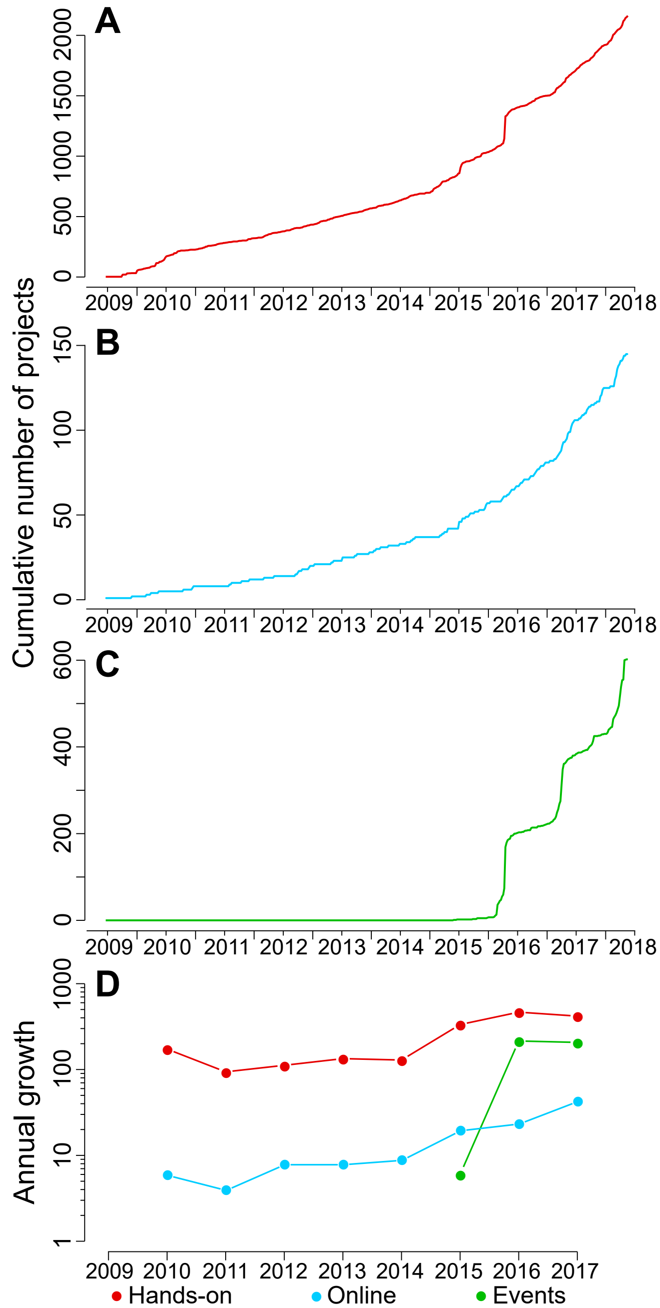

Fig S1. Growth in the number of citizen science projects on the clearinghouse platform SciStarter and/or on the online platform Zooniverse, as a function of project type. For Scistarter, only projects that are vetted against harm, crowdfunding, junk science, or are strictly education/outreach are accepted. (A) Hands-on projects defined by some-to-all of the project requires a participant to conduct a task(s) other than online (SciStarter). (B) Projects conducted entirely online (Zooniverse). (C) One-off events, including but not limited to blitzes, trainings, meet-ups and in-person transcription events (SciStarter). (D) Annual growth as a function of project type.

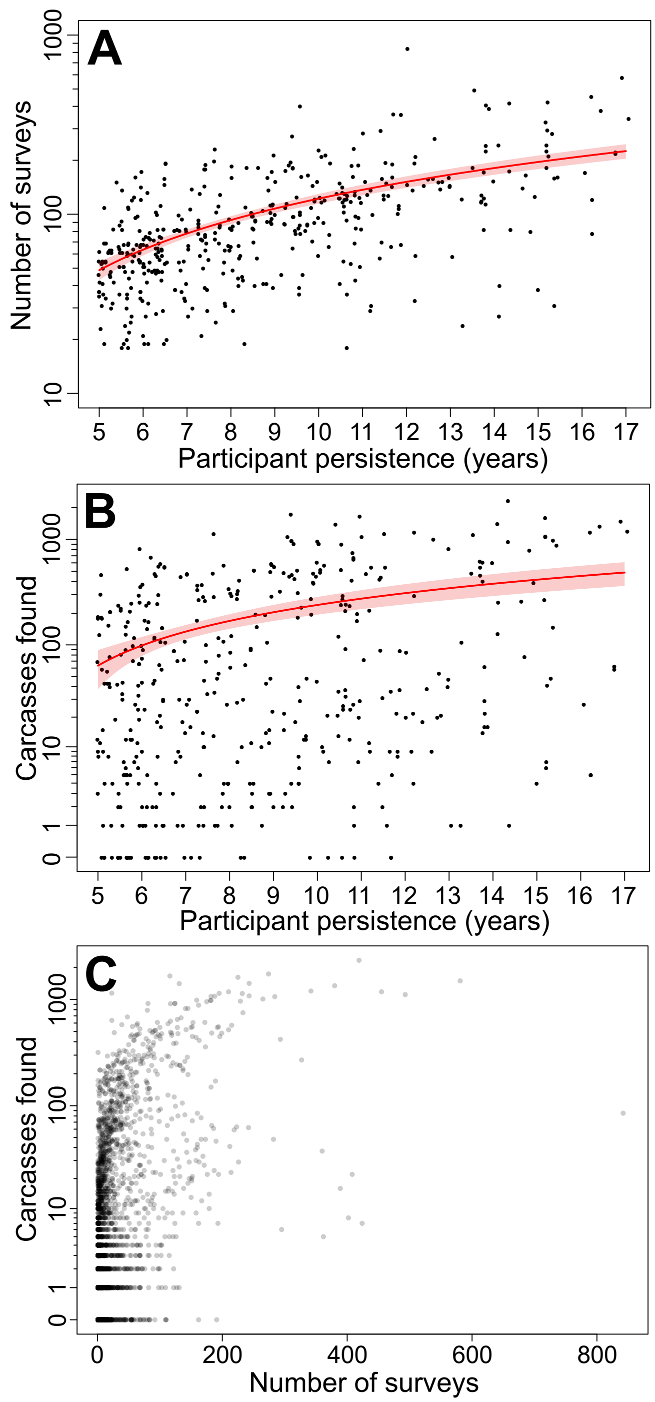

Fig. S2. The relationship between participant effort and persistence (mean and 95%CI) for long-term participants (duration >5 years; N = 491) modeled as linear relationship. (A) Effort measured as the number of (monthly) surveys performed, with fitted model ($N_{surveys}$ ~ 14.6*Duration in years) overlaid. (B) Effort measured as the number of carcasses found, with fitted model (N_{birds} ~ 35.2*Duration in years) overlaid. (C) Number of carcasses found as a function of the number of surveys performed for all participants (N=3,286). Dots are participants.

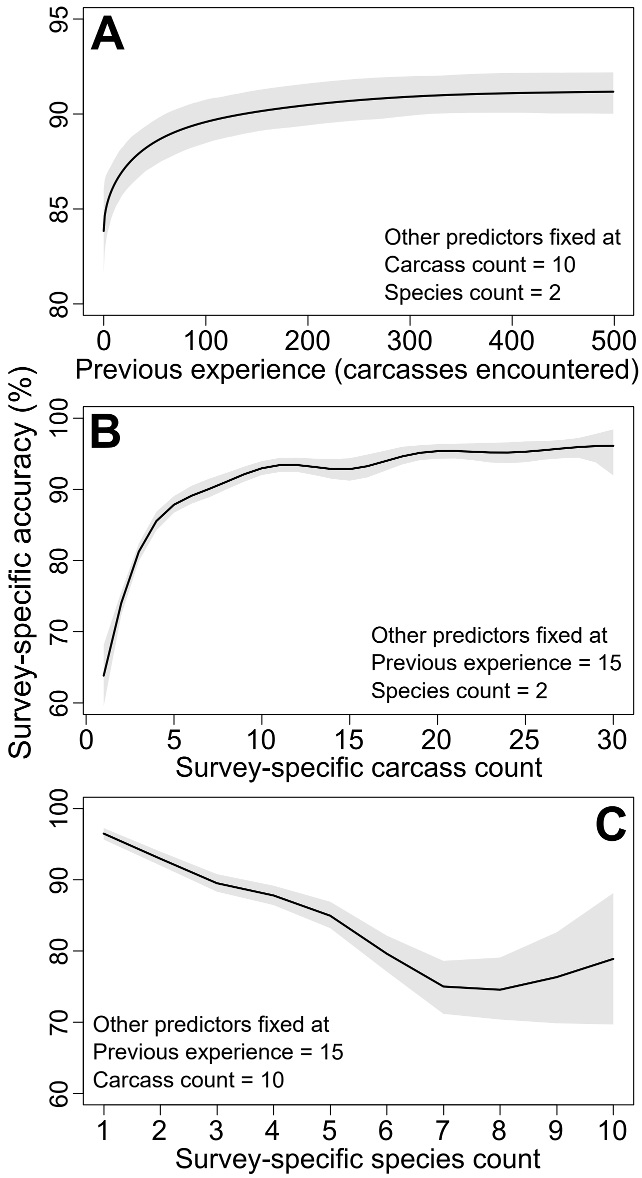

Fig. S3. Survey-specific accuracy (mean and 95% CI) modeled using generalized additive models (GAMs) with smooth terms of (A) previous beached bird experience, (B) survey-specific carcass abundance, and (C) survey-specific species richness.

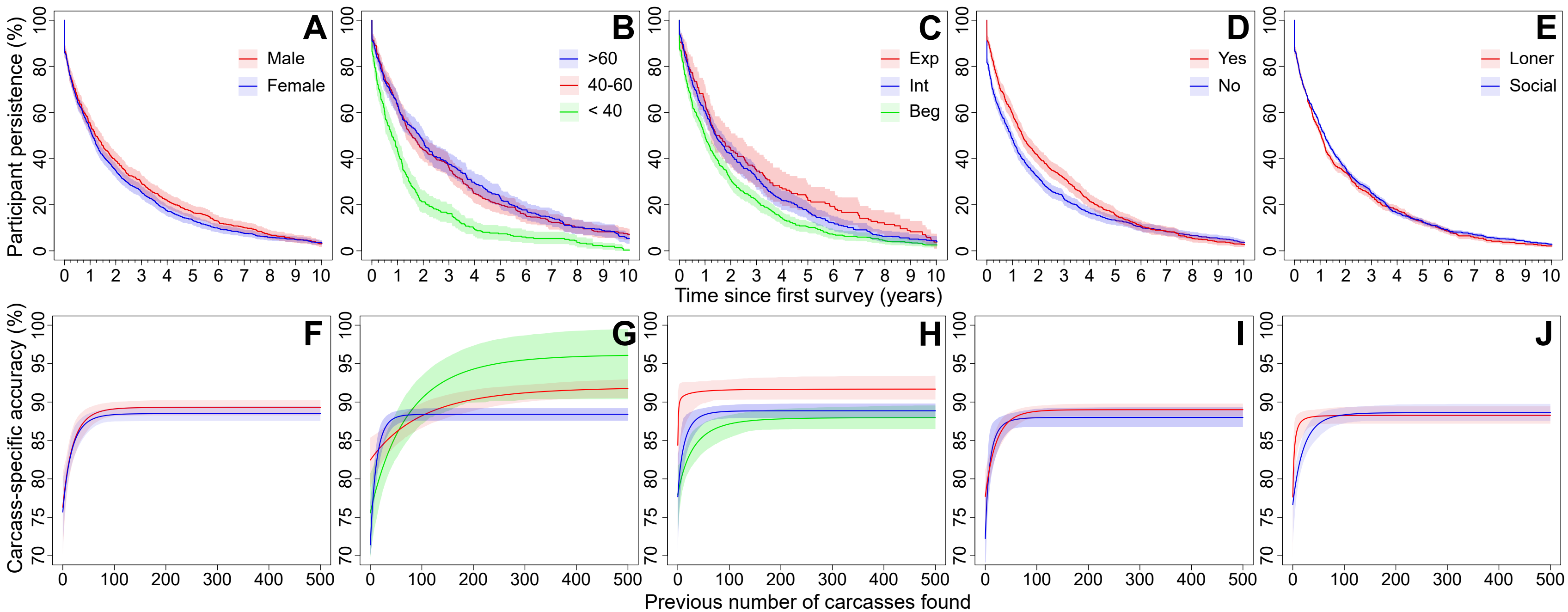

Fig. S4. Retention (A-E) and accuracy (F-J) modeled as in Figure 2A and 2C, as a function of demographic and interest categories: (A, F) gender; (B, G) age at training; (C, H) bird identification expertise; (D, I) previous experience in citizen science; (E, J) sociality.

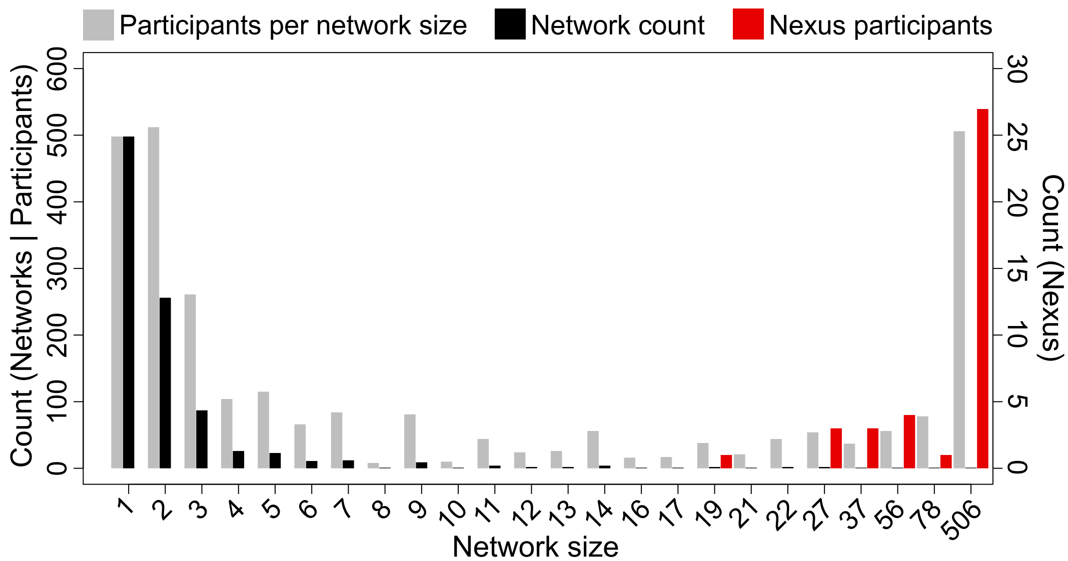

Fig. S5. Frequency distribution of network size (black), cumulative participants at size (grey), and distribution of nexus people (N = 39) across COASST social networks (red).

Table S1. List of citizen science projects and data sources for all information in the retention landscape (Fig. 1) ordered alphabetically by source within project category (e.g., hands-on deductive data). We used Google Scholar and Web of Science to find program retention statistics with search terms “retention, or turnover, or contribution, or engagement" and "citizen scien*, or public participation in scientific research" or "*blitz." Because this effort predominantly returned articles featuring online programs, we also contacted select hands-on citizen science project coordinators for project statistics (indicated in parentheses). See acknowledgements for direct data sources (e.g., without references).

Project Name	**Citation**
Hands-on deductive data	
University of Minnesota Extension Forest Pest First Detector	(4)
River Herring Monitoring program	(5)
Virginia Master Naturalist	(6)
FreshWater Watch - Adopt a River	(7)
Great Pollinator Project	(8)
Reef Check Foundation’s California Program (RCCA)	(9)
Invaders of Texas	(10)
Plants of Concern	(11)
Pollination Investigators	(12)
Tucson Bird Count	(13)
Wabash Sampling Blitz	(14)
Monarch Larvae Monitoring Project	(15)
BTO/JNCC/RSPB Breeding Bird Survey	(16)
Waterways Breeding Bird Survey	
Garden Birdwatch	
Living with Mammals	
BeeWatch	(17)
Washington Sea Grant Crab Team (CT; dabblers: 3.3%, 1y+: 86.2%)	
Great Sunflower Project (GSP; dabblers: 29%, 1y+:71%)	
Puget Sound Seabird Survey (PSSS; dabblers: 0%, 1y+: 65.2%)	
Reef Environmental Education Foundation (REEF; dabblers: 38.4%, 1y+: 20.8%)	
Nature's Notebook (NN; dabblers: 27.3%, 1y+: 12.9%)	
FeederWatch (FW; dabblers: 5%)	
COASST (this study)	
Hands-on sampling	
Greenspace Information for Greater London Community Interest Company (GiGL)	(18)
iRecord	
iSpot	
Community Collaborative Rain, Hail and Snow Network (CoCoRaHS)	(19)
Online classification	
Season Spotter	(20)
Happy Match	(21)
Galaxy Zoo Hubble	(22)
Galaxy Zoo Supernovae	
Old Weather	
Moon Zoo	
Solar Stormwatch	

Bat Detective	(23)
Condor Watch	
Cyclone center	
Disk Detective	
Galaxy Zoo	
Milky Way Project	
Notes from Nature	
Operation War Diary	
Planet Four	
Planet Hunters	
Plankton portal	
Radio Galaxy Zoo	
Seafloor Explorer	
Snapshot Serengeti	
Sunspotter	
Worm Watch Lab	
Online gaming	
Phylo	(24)
Forgotten Island	(21)
Fraxinus	(25)

Table S2. Summary of Generalised Additive Models fitted to survey-specific accuracy data. Presented statistics are akaike information criterion (AIC), % of deviance explained D^2, Δ_{AIC} = AIC – min(AIC) and akaike weight, $w = e^{-\Delta_{AIC}/2}$.

Rank	Predictors	AIC	D^2	Δ_{AIC}	w_{AIC}
1	$s(N_{bird}) + s(N_{species}) + s(Experience)$	8972.8	10.1	0	1
2	$s(N_{bird}) + s(N_{species})$	9063.6	8.8	90.8	<1e-16
3	$s(N_{bird}) + s(Experience)$	9340.2	4.9	367.4	<1e-16
4	$s(N_{bird})$	9411.5	3.8	438.7	<1e-16
5	$s(N_{species}) + s(Experience)$	9525.9	2.5	553.1	<1e-16
6	s(Experience)	9556.9	1.9	584.1	<1e-16
7	$s(N_{species})$	9653.6	0.6	680.8	<1e-16
8	Null	9688.3	0	715.5	<1e-16

References

1. Stan Development Team (2016). RStan: the R interface to Stan. R package version 2.14.1. http://mc-stan.org/.
2. Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J Royal Statis Soc (B)* 73:3-36
3. Csardi, G, Nepusz, T (2006). The igraph software package for complex network research, InterJournal, Complex Systems 1695. http://igraph.org
4. Andow DA, Borgida E, Hurley TM, Williams AL (2016) Recruitment and Retention of Volunteers in a Citizen Science Network to Detect Invasive Species on Private Lands. *Environ Manag* 58:606-618.
5. Bieluch KH, Willis T, Smith J, Wilson KA (2017) The Complexities of Counting Fish: Engaging Citizen Scientists in Fish Monitoring. *Maine Pol Rev* 26:9-18.
6. Blanton B, Crall AW, Prysby MD, Mellor DT, Brown T (2014) Virginia Master Naturalist Strategic Planning Report, 2015-2020. Virginia Master Naturalist Program: Charlottesville, VA. 161p.
7. Cunha DG, et al. (2017) Citizen science participation in research in the environmental sciences: key factors related to projects' success and longevity. *Anais da Academia Brasileira de Ciências, 89*(3), 2229-2245.
8. Domroese MC, Johnson EA (2017) Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biol Conserv* 208:40-47.
9. Freiwald J, Wehrenberg M, Wisniewski C, Hodgson G (2013) Status of Rocky Reef Ecosystems in California 2006 – 2011. Reef Check Foundation, Pacific Palisades, CA, USA.
10. Gallo T, Waitt D (2011) Creating a successful citizen science model to detect and report invasive species. *BioScience* 61:459-465.
11. Havens K, Vitt P, Masi S (2012) Citizen science on a local scale: the Plants of Concern program. *Front Ecol Environ* 10:321-323.
12. Kleinke B, et al. (2018) Identifying Barriers to Citizen Scientist Retention When Measuring Pollination Services. *Cit Sci Theor Prac* 3:2
13. McCaffrey RE (2005) Using citizen science in urban bird studies. *Urban Hab* 3:70-86.
14. Muenich R, et al. (2016) The Wabash Sampling Blitz: A study on the effectiveness of citizen science. *Cit Sci Theor Prac* 1:3
15. Prysby MD, Oberhauser KS (2004) Temporal and geographic variation in monarch densities: citizen scientists document monarch population patterns. *The Monarch Butterfly: Biology and*

Conservation, eds Oberhauser KS and Solensky MJ (Cornell University Press, Ithaca, New York, USA), pp. 9-20.

16. Roos S, Johnston A, Noble D (2012) UK hedgehog datasets and their potential for long-term monitoring. *BTO Research Report* 598.
17. Siddharthan A, et al. (2016) Crowdsourcing without a crowd: Reliable online species identification using Bayesian models to minimize crowd size. *ACM Transactions on Intelligent Systems and Technology (TIST), 7*:45.
18. Boakes EH, et al. (2016) Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Scientific reports, 6*, 33051.
19. Sheppard S A, Turner J, Thebault-spieker J, Zhu H, Terveen L (2017) Never too old, cold or dry to watch the sky: A survival analysis of citizen science volunteerism. *Proceedings of the ACM on Human-Computer Interaction* 1:94:1–94:21, 2017.
20. Crall A, et al. (2017) Volunteer recruitment and retention in online citizen science projects using marketing strategies: lessons from Season Spotter. *J Sci Comm* 16:A01:1-29
21. Prestopnik N, Crowston K, Wang J (2017) Gamers, citizen scientists, and data: Exploring participant contributions in two games with a purpose. *Comp Human Behav* 68:254-268.
22. Sauermann H, Franzoni C (2015) Crowd science user contribution patterns and their implications. *Proc Natl Acad Sci U.S.A.* 112:679-684.
23. Segal A, et al. (2015) Improving productivity in citizen science through controlled intervention. *Proceedings of the 24th International Conference on World Wide Web* (ACM, New York). pp. 331-337. Florence, Italy. May 18–22, 2015.
24. Kawrykow A, Ret al. (2012) Phylo: a citizen science approach for improving multiple sequence alignment. *PloS ONE 7*:e31362.
25. Rallapalli G, et al. (2015) Lessons from Fraxinus, a crowd-sourced citizen science game in genomics. *eLife* 4:e07460