OXFORD

Experimental evidence that behavioral nudges in citizen science projects can improve biodiversity data

Corey T. Callaghan ⓘD, Maureen Thompson, Adam Woods, Alistair G. B. Poore, Diana E. Bowler, Fabrice Samonte, Jodi J. L. Rowley, Nadiah Roslan, Richard T. Kingsford, William K. Cornwell and Richard E. Major

Corey T. Callaghan (callaghan.corey.t@gmail.com) and Diana E. Bowler are affiliated with the German Centre for Integrative Biodiversity Research (iDiv) Halle–Jena–Leipzig, in Leipzig, Germany. Corey T. Callaghan, Maureen Thompson, Fabrice Samonte, Jodi J. L. Rowley, Richard T. Kingsford, and William K. Cornwell are affiliated with the Centre for Ecosystem Science in the School of Biological, Earth, and Environmental Sciences at the University of New South Wales Sydney, in Sydney, New South Wales, Australia. Corey T. Callaghan is also affiliated with the Institute of Biology at Martin Luther University Halle–Wittenberg, in Halle (Saale), Germany, and with the Department of Wildlife Ecology and Conservation at the Fort Lauderdale Research and Education Center, University of Florida, in Davie, Florida, in the United States. Maureen Thompson, Adam Woods, Jodi J. L. Rowley, Nadiah Roslan, and Richard E. Major are affiliated with the Australian Museum Research Institute, at the Australian Museum, in Sydney, New South Wales, Australia. Alistair G. B. Poore and William K. Cornwell are affiliated with the Ecology and Evolution Research Centre, in the School of Biological, Earth, and Environmental Sciences at the University of New South Wales, Sydney, in Sydney, New South Wales, Australia. Diana E. Bowler is also affiliated with the UK Centre for Ecology and Hydrology, in Wallingford, England, in the United Kingdom.

Abstract

One way to improve the value of citizen science data for a specific aim is through promoting adaptive sampling, where the marginal value of a citizen science observation is dependent on existing data collected to address a specific question. Adaptive sampling could increase sampling at places or times—using a dynamic and updateable framework—where data are expected to be most informative for a given ecological question or conservation goal. We used an experimental approach to test whether the participants in a popular Australian citizen science project—FrogID—would follow an adaptive sampling protocol aiming to maximize understanding of frog diversity. After a year, our results demonstrated that these citizen science participants were willing to adopt an adaptive sampling protocol, improving the sampling of biodiversity consistent with a specific aim. Such adaptive sampling can increase the value of citizen science data for biodiversity research and open up new avenues for citizen science project design.

Keywords: citizen science, community science, biodiversity monitoring, species richness, FrogID

Biodiversity data collected through citizen science, also referred to as *community science*, are rapidly becoming the predominant source of biodiversity data across the world (Chandler et al. 2017), accounting for at least 80% of all data available since 2010 (www.gbif.org; supplemental text S1) in the largest biodiversity aggregator, the Global Biodiversity Information Facility. Global citizen science initiatives such as iNaturalist accumulate an average of approximately 80,000 species observations per day across the world (www.inaturalist.org; supplemental text S2). Such data are expanding our understanding of biodiversity patterns in space and time and informing conservation policy and practice (e.g., Schuster et al. 2019, Billaud et al. 2021, Forister et al. 2021, Kirchoff et al. 2021). But paradoxically, more data do not necessarily mean increased knowledge about species' distributions; new observations are often collected in already sampled places, creating redundancies underlying many sources of biodiversity data (Boakes et al. 2010, Courter et al. 2013, Tiago et al. 2017a).

Citizen science initiatives vary in scope, design, and structure (Haklay 2013, Pocock et al. 2017), influencing the extent of biases in the associated data (Welvaert and Caley 2016). Citizen science initiatives fall along a continuum (e.g., Pocock et al. 2015, Welvaert and Caley 2016), ranging from unstructured recording (e.g., in which little training is needed to participate and contribute opportunistic or incidental observations and with which few metadata are associated) to focused recording (e.g., with minimal workflows and guidelines but increased metadata collected with each observation such as search effort) to structured recording (e.g., prescribed sampling in space and time by mostly trained and experienced volunteers, usually but not always with metadata and with site selection included in the design). Because of the ease with which volunteers can participate, unstructured data projects generally provide data at the largest spatial and temporal scales, with minimal metadata (e.g., the date of observation and the location of the observation), but these data often contain the most bias, especially spatial bias (Geldmann et al. 2016). Although various statistical methods can account for noise and biases in citizen science data (e.g., Bird et al. 2014, Isaac et al. 2014, Johnston et al. 2020), the value of the data could be more directly improved by changing how it is collected.

One way to improve biodiversity data is to guide citizen science sampling—that is, to add structure to the sampling process (Pocock et al. 2015, Callaghan et al. 2019a, 2019b). For example, areas where no observations have been reported could be prioritized for future sampling, filling in our understanding of biodiversity (Fontaine et al. 2021). In the present article, we call this *adaptive sampling* (*sensu* Zeng and Xiang 2017, Takahashi et al. 2022). The marginal value of a potentially new citizen science observation (i.e., the relative value of a given observation relevant to other possible observations) is dependent on the temporal or spatial attributes of recent citizen science observations submitted to the platform and on the question (i.e., statistical outcome) of the proposed objective. Adaptive sampling can inform sampling at places or times where the data are expected to be most informative for a given question in ecology or conservation, where such questions

could include assessments of species temporal trends, spatial patterns, or even community properties. Adaptive sampling in citizen science projects is in its infancy, with multiple studies looking at different optimal sampling designs given a particular goal or outcome of a citizen science project (e.g., Callaghan et al. 2019a, Kays et al. 2021).

The success of an adaptive sampling scheme depends in part on appealing to the motivations of the participants to change their recording behavior. This notion is derived from the theory of behavioral nudging—the concept of influencing the motives and incentives of groups or individuals through positive reinforcement or indirect suggestions (Thaler and Sunstein 2008). Such dependence on understanding individuals' motivations illustrates the importance of understanding the different intrinsic and extrinsic motivations of citizen science participants (Pateman et al. 2021) and potential interactions (West et al. 2021). Participant motivations are complex, and in the present article, we define *intrinsic motivations* as those through which the participants find an activity inherently interesting or satisfying and *extrinsic motivations* as those through which the participants work to gain an instrumental or external goal or reward (see West et al. 2021). Just as the particular goals or outcomes of citizen science projects are diverse, so too are the citizen science participants—diverse in terms of their data contribution and motivations (August et al. 2020, Pateman et al. 2021, West et al. 2021). Generally, volunteer participation in citizen science programs (Anđelković et al. 2022) is motivated by an intrinsic willingness to contribute to conservation or environmental concerns (Tiago et al. 2017b, Larson et al. 2020, West et al. 2021), social or competitive features of a project (Eveleigh et al. 2014, Pateman et al. 2021), or personal reasons (West et al. 2021, Anđelković et al. 2022). In this instance, a nudge could be represented by conveying the information about the importance of a given citizen science observation for research or conservation. For example, if the benefits of data collection for conservation purposes are clearly articulated, then the participants motivated by conservation (West et al. 2021, Angello et al. 2022) may be more likely to adopt sampling nudges. Alternatively, the participants interested in bettering themselves (i.e., participating to learn something or further one's career; West et al. 2021) may be willing to adopt an adaptive sampling protocol if it helps that individual achieve that goal.

In contrast, some citizen science participants are motivated by competition (Bowser et al. 2013), sometimes in addition to or complementing other intrinsic motivations. In this instance a nudge—going beyond conveying the importance of a given observation for research—could include a gamified aspect expected to result in uptake of an adaptive sampling protocol. Gamification has proven successful across fields—for example, by increasing health-related behavior, academic performance, and environmental sustainability (Morford et al. 2014, Manzano-León et al. 2021). Self-presentation, self-efficacy, social bonds, and playfulness may all serve as different motivations that lead to the success of gamification in crowdsourcing (Feng et al. 2018) and are therefore important to fully understand the potential of gamification to produce successful behavioral nudges. In citizen science initiatives, previous work has shown the potential for success of gamification (e.g., Xue et al. 2016), predominantly through the use of leaderboards (Wood et al. 2011), used to encourage sampling among the participants. But such leaderboards are generally focused on the number of species or the number of observations—neither of which necessarily increases knowledge of biodiversity or improves decision-making for conservation (Bayraktarov et al. 2019). This contrasts with a leaderboard that quantifies the collective value of a participant's observations.

We are unaware of any formal quantification of the extent to which citizen science participants are willing to respond to an adaptive sampling scheme and whether behavioral nudges can be used to encourage different sampling of biodiversity. We used an experimental approach to test whether the participants in a popular Australian citizen science project—FrogID—would follow an adaptive sampling protocol aimed at maximizing the understanding of frog diversity. We had three experimental groups: one presented with a dynamic map of optimal sampling locations updated biweekly that communicated to the participants where their data collection would be most valuable (i.e., on the basis of the potential to add to our understanding of frog diversity), another presented with the same dynamic map of optimal sampling locations but with a leaderboard highlighting each participant's cumulative total of valuable observations in the region, and a third a priori control group who were not shown any maps.

Using this experimental design, we tested the following hypotheses. First, we hypothesized that behavioral nudges would have a positive impact on objectively better sampling of biodiversity, as would be judged by the participants' proportionally submitting higher priority samples. Second, we hypothesized that a leaderboard would have a further impact on higher priority sampling such that the participants in the leaderboard experimental group would submit proportionally higher priority samples. And third, we hypothesized that if the participants were adapting the behavioral nudges, then we would be able to detect this change in the difference in spatial bias of the samples. Our results have wide-reaching implications for the use of adaptive sampling in current and future biodiversity-related citizen science initiatives.

An experimental approach
FrogID citizen science project

For our experiment, we used the FrogID citizen science platform. FrogID (Rowley et al. 2019, Rowley and Callaghan 2020) is a citizen science project led by the Australian Museum, in Sydney, Australia (see https://australian.museum), launched in November 2017 and aimed at gathering data on Australian frogs. The volunteers submit 20–60 acoustic recordings of calling frogs, which are then subsequently identified by a team of experts at the Australian Museum. The recordings are made via a smartphone app and are geolocated to a single point, with associated accuracy estimation. The data are in the form of presence-only observations. To date, the FrogID project has more than 750,000 observations of frogs submitted by more than 30,000 participants. Each frog observation in the data set is hereafter referred to as a *sample*.

Experimental design

We chose six study regions defined by local government areas (hereafter, *study regions*) for our experimental design (supplemental figure S1). The study regions were selected a priori on the basis of relatively equal-size areas, a reasonable number of active FrogID users, a diverse range of habitats, and the known level of frog diversity. The six study regions in our project were assigned to one of three treatments such that there was no systematic bias among treatments in terms of these aforementioned characteristics (see figure S1): the dynamic map group (Central Coast and Wollongong), the dynamic map and leaderboard group (Hornsby and Blue Mountains), and the control group (Wingecarribee and Lake Macquarie). All maps were presented via a website deployed using html and the react-leaflet application. The dynamic map group was presented with a map illustrating the area of the study region and how
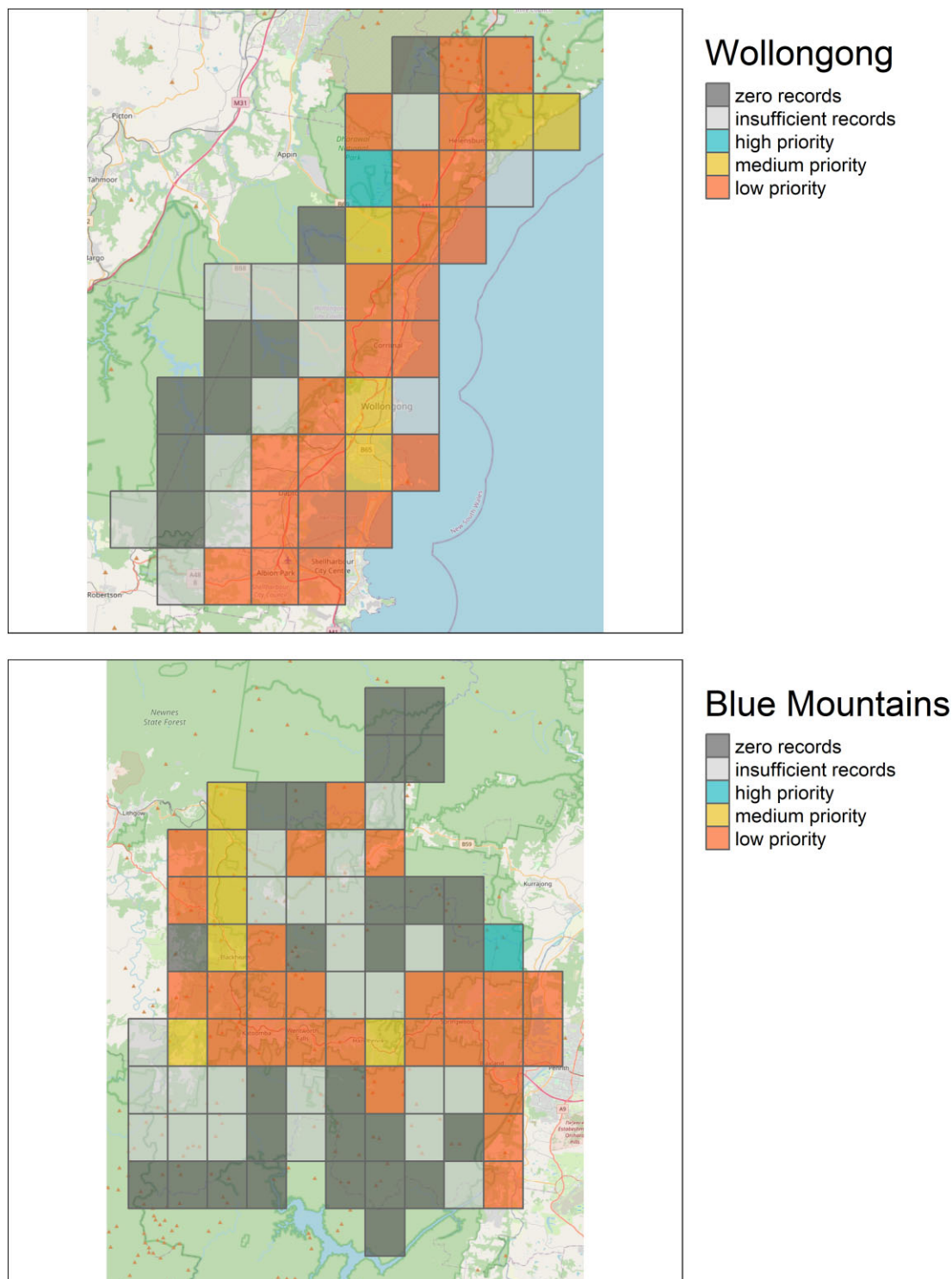
Figure 1. An example of the sampling priority map shown to the participants of the project, summarizing our workflow to qualitatively illustrate the sampling priority of the areas throughout the study regions. Users were shown an interactive version that could be zoomed in (see supplemental videos S1 and S2 for a better overview of the website).

the associated priority (see below) varied throughout the study region (e.g., figure 1). The dynamic map and leaderboard group was presented with the same map, calculated as in the dynamic map group, but they were also presented with a leaderboard. The leaderboard ordered FrogID registered usernames on the basis of the score of the observations from the dynamic maps. The score was calculated by assigning a point value to each of

the five grid categories: low priority, 1; medium priority, 2; high priority, 3; insufficient records, 4; and zero records, 5 (see below for details). And for every observation a user submitted in the grid cell, there were up to a maximum of five observations on any given day. The values were then multiplied and added and were provided to the users via the website (see supplemental videos S1 and S2). The control group represented two study regions, chosen

a priori as controls to account for the increasing rise in FrogID users and contributions (Rowley et al. 2019) during the study (i.e., distinguishing increased records on the basis of the natural growth of the project and as a response to our study).

Deriving a dynamic map of priority locations

We chose estimating species richness as our scientific objective, given its fundamental importance in understanding biodiversity and prioritizing conservation efforts (Yocoz et al. 2001) and the potential to estimate species richness using citizen science data (Callaghan et al. 2020). We overlaid 0.05 by 0.05–degree grid cells over each study region (supplemental figure S2), producing biweekly dynamic maps throughout a year (e.g., figure 1). We chose this spatial resolution on the basis of our ability to aggregate FrogID observations into spatial units that could estimate species richness (i.e., we wanted to collate a minimum number of observations that is more likely at larger spatial scales). Also, this resolution was chosen because most FrogID users contribute data within 1 kilometer of their home, but we wanted to test the ability of adaptive sampling, requiring a spatial resolution larger than approximately 1 kilometer. The grid cells were categorized into sampling priority statuses (a qualitative representation of the importance of a sample from a given grid cell to aid our understanding of frog diversity): zero records, insufficient records, high priority, medium priority, and low priority. *Zero records* was any grid cell with no FrogID submissions. *Insufficient records* applied to grid cells rarely sampled; we used a cutoff of 10 FrogID submissions, because this was the minimum number producing reliable estimates of species richness (see supplemental figure S3). A minimum number was necessary for the statistical analysis to estimate completeness of a grid cell (see below for details); for less than 10 observations, it was more likely that only one species was recorded, which would produce unreliable species richness estimates. Different cutoffs (for insufficient records) could be used, but we chose 10 FrogID submissions because too few grid cells were classified to provide a robust estimate of species richness in a grid cell. Regardless of how the insufficient records threshold is defined, our results of differential sampling among high, medium, and low priority cells would still be robust. Moreover, the reason for few records in both cells classified as *insufficient records* and *zero records* could be many, including a lack of frogs within that specific grid cell (i.e., little frog habitat) or inaccessible habitat (i.e., predominantly private lands), which is why we separated these categories from the others (high, medium, and low priority cells). To categorize a grid cell as high, medium, or low priority, we used an estimate of completeness with respect to the observed and expected species richness in a grid cell. This was estimated using the iNEXT package in R (Chao et al. 2014, Hsieh et al. 2016), which takes a sampling assemblage of abundance N with species richness i, and calculates a sample completeness curve with respect to the sample size. This curve is an aggregate of the interpolated and extrapolated species accumulation and the estimated asymptote, along with a confidence interval, for species richness. Inputs into the iNEXT function were generated by obtaining abundance data (i.e., presence-only counts of each species) for each spatial grid cell in a given area. First, we estimated species richness on the basis of the recordings submitted in a grid and then divided the observed species richness by the estimated species richness (i.e., a grid where the species richness was well estimated would be equal to 1 according to iNEXT). The inverse of these completeness values (i.e., incompleteness) was used to assign sampling priority status to a grid cell. For example, a grid cell with a completeness

value of 1 was classified as a low priority cell, because sampled richness was already likely to be complete or close to complete (see supplemental figure S4 for an illustration). For values greater than 1, we categorized these grid cells—those with more than 10 records—into high (the top 33%), medium (the middle 33%), and low (the bottom 33%) priority by dividing the range of values into three, for qualitative representation in our dynamic maps. For example, if the inverse completeness value was at least two-thirds of the maximum value of completeness values, then the grid cell was assigned high priority.

Because the estimates of species richness change and sampling completeness changes with new observations, this procedure was updated biweekly, adding new validated records from the previous period. We refer to each 2-week period as a sampling period, and our study therefore included a total of 26 sampling periods. Importantly, in our framework, a grid cell was dependent on the other grid cells in that study region, and consequently, a grid cell could be low priority one time period but subsequently high priority.

Recruitment of participants

All of the participants who had submitted records from the four experimental study regions before the study began were initially invited to join the project. The participants were emailed and told of our pilot study to encourage better sampling of frogs (supplemental figure S5). A FrogID user could receive more than one recruitment email if they submitted an observation in more than one of the experimental study regions. The users were given the option to opt out of email communication. However, because participants could join the FrogID project throughout the entirety of our study, we reanalyzed the user data to identify any potential new FrogID participants contacted during the next update email (e.g., supplemental figure S6). In total, we emailed the participants four times: at the start of the project, 3 months after the start, 6 months after the start, and 9 months after the start. After the first email, we quantified the number of website hits to track the usefulness of sending the recruitment emails, which produced significant spikes in website visits when these emails were sent (supplemental figure S7).

Statistical analysis

For statistical analyses, the control group was treated as the comparison, representing an approximation of overall change without dynamic maps presented to the FrogID users. To empirically compare the relative efforts made toward incomplete cells between control (Wingecarribee and Lake Macquarie) and experimental groups (Central Coast, Blue Mountains, Wollongong, Hornsby), the number of samples per cell for each priority status was calculated for each study region. Rather than comparing the raw numbers of samples, we standardized for the spatial area of each sampling priority status. To obtain the number of samples per square, the ratio of the total number of submitted samples and the total number of cells generated across all sampling periods was calculated for each respective sampling priority status. This process was repeated across each of the 26 sampling periods of the experiment. For example, in the Blue Mountains study region, there were 28 spatial grid cells classified as high priority sampling status, with a total of 451 frog records submitted within those grid cells. This led to 16.1 samples per cell in the areas classified with a high priority sampling status.

To statistically test for an effect of experimental treatments, we asked two questions addressing our first two hypotheses from the introduction: Was there a difference between the

experimental groups (i.e., all four experimental study regions presented with a dynamic map) and the control group (i.e., the two control study regions)? And was there a difference between the two experimental groups with and without a leaderboard? We used a generalized linear mixed effects model with a Poisson error distribution, where our response variable was the number of samples with sampling priority status as a predictor variable, as well as an interaction term for sampling priority status by experimental group. We used a random effect for study region to account for potential differences among study regions, with an offset term for the number of cells (log-transformed) in each sampling priority status to account for differences in sampling areas. We defined the Wald contrasts of the model so that the interaction regression coefficients and their associated p-values would test the two questions defined above.

We further tested for effects of our dynamic maps by assessing spatial bias in sampling, comparing whether the observed bias matched the expected bias if the participants followed the advice of our dynamic maps. This analysis corresponded to our hypothesis that we would be able to detect the effect of behavioral nudges in the spatial bias of the sampling. We compared the spatial bias of FrogID samples across the experimental period (October 2020–October 2021) with the period of the prior year (October 2019–October 2020). A simple measure of spatial bias in sampling (e.g., spatial bias before and spatial bias after) was not informative, because some bias in sampling locations was anticipated because species richness is heterogeneous. Therefore, we generated a null model of what spatial bias would look like if people followed the dynamic maps we provided. This was generated by distributing n random samples on the basis of the proportion of cells of each priority status. For example, if 10 samples were made within the spatial map consisting of five low priority cells, three medium priority cells, and two low priority cells, then the null model was generated by randomly sampling 50% of the samples within low priority areas, 30% within medium priority areas, and 20% in high priority areas. To account for the differing numbers of samples between experimental and control periods, two equilibrium models were created. The spatial bias was determined by calculating the average mean squared distances of all point locations in the set of FrogID records, and the average mean squared distances of all point locations in the set of model random samples (i.e., the null model). This process was repeated for all 26 sampling periods, to produce distributions of delta average mean squared distances, separated by the experimental and control time frames specified. This was further processed for all six chosen study regions involved in the study. For our control groups, we generated dynamic priority maps, although we note that no dynamic maps were presented to the observers in these study regions; we used these as our null model of spatial sampling. To statistically test this, we followed the procedure described above but used a linear mixed effects model with a Gaussian error distribution, where our response variable was the delta mean squared distance with the period (before or after the experiment) as a predictor variable, as well as an interaction term for the period (before or after the experiment) and the experimental group. We used a random effect for study region to account for potential differences among study regions. Similarly, we defined the Wald contrasts of the model so that the interaction regression coefficients and their associated p-values would test the two questions defined above.

Data accessibility

Not all of the raw FrogID data can be made fully available because the data contain identifiable information for the FrogID users. However, the processed and summarized data used to reproduce the figures and analyses are available in a Zenodo repository here: https://doi.org/10.5281/zenodo.7589542.

Findings

Across our four experimental study regions (Central Coast and Wollongong were the dynamic map treatment, and Hornsby and Blue Mountains were the dynamic map and leaderboard treatment), a total of 38,732 recordings were submitted during our 1-year study period from 1710 FrogID users (supplemental figure S8). The Central Coast had the most submissions ($n = 21,780$), followed by the Blue Mountains ($n = 7466$), Wollongong ($n = 5484$), and Hornsby ($n = 3912$). These submissions corresponded to 38, 21, 21, and 15 frog species, respectively, within each study region. In contrast, our control study regions had 3131 submissions from Lake Macquarie, corresponding to 29 frog species and 2171 submissions from Wingecarribee corresponding to 21 frog species.

We found empirical differences in the sampling patterns between our experimental groups (i.e., those study regions presented with a dynamic map) and our control groups (i.e., those study regions not presented with a dynamic map) on the sampling priority status (figure 2). Relative to the available area in a study region, the low priority cells in the control study regions accounted for 36% (Lake Macquarie) and 61% (Wingecarribee) of the sampling, whereas for our study regions presented with a dynamic map low priority cells accounted for only 12% (Wollongong) and 23% (Central Coast) of the sampling, and for study regions presented with a dynamic map and a leaderboard, low priority cells accounted for only 15% (Blue Mountains) and 22% (Hornsby) of the sampling. Furthermore, the high priority cells in the control study regions accounted for 32% (Lake Macquarie) and 15% (Wingecarribee) of the sampling, whereas for our study regions presented with a dynamic map, high priority cells accounted for 58% (Wollongong) and 8% (Central Coast) of the sampling, and for the study regions presented with a dynamic map and a leaderboard, high priority cells accounted for 73% (Hornsby) and 67% (Blue Mountains) of the sampling.

We found a statistically significant difference between the sampling of priority status cells in the experimental regions (i.e., the dynamic map or the dynamic map and leaderboard study regions combined) and the control study regions. This statistical effect was strongest for high priority cells, followed by medium priority and low priority cells (figure 3). In contrast, there were no statistically significant differences in sampling cells with insufficient records ($n = 220$ total cells across the study period) or zero records ($n = 74$ total cells across the study period). Therefore, our experiment increased sampling in high priority areas but not at the cost of other areas; we found an effect of increased sampling in high, medium, and low priority areas. Also, we found differences between our different experimental groups, but the effects were not consistent. The study regions presented with a leaderboard had a statistically significant effect of more sampling in high priority cells, but the converse was true for medium priority cells. And we found no statistically significant difference between our two experimental study regions for low priority status sampling or insufficient records status, but there was more sampling for zero records status for the group with a leaderboard, albeit this was a relatively small number of total samples during the experiment.

In further support of the ability to shift citizen science sampling through an adaptive sampling scheme, we found that the patterns in spatial bias more closely matched the guidance from the dynamic maps for the experimental study regions than for the control study regions (figure 4, supplemental figure S9). These
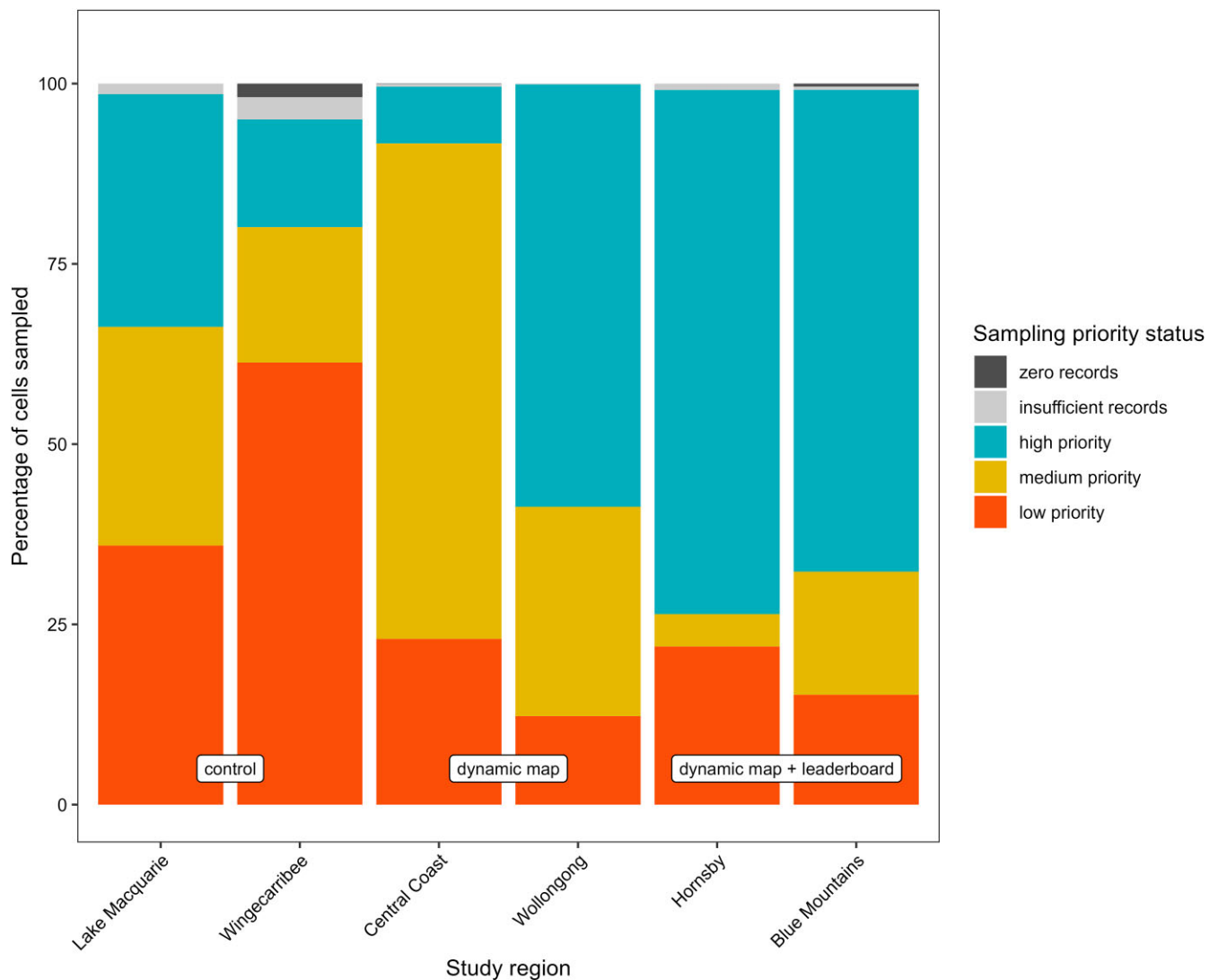
Figure 2. The relative sampling throughout the study period, represented by the percentage of samples standardized by the area of priority status, stratified by study region. Lake Macquarie and Wingecarribee were control study regions, whereas the participants in Central Coast and Wollongong were presented with a dynamic map and those in Hornsby and Blue Mountains were presented with a dynamic map and a leaderboard.

patterns were most striking for the two study regions presented with a dynamic map where the spatial bias compared with the null was less after the experiment compared with before the experiment. For the control study regions, the opposite pattern was found. For the two study regions with a dynamic map and a leaderboard, the spatial bias and the null both before and after the experiment were comparable.

The potential of behavioral nudges for future citizen science sampling

Using experimental evidence, we demonstrated for the first time that citizen science participants are willing to adopt new data collection strategies that improve sampling of biodiversity, consistent with a specific aim. The participants in the FrogID project in Australia followed behavioral nudges, presented through the use of dynamic maps, which led to an increase in sampling of high priority and medium priority cells when compared with study regions in which the participants were not presented with dynamic maps (figures 2 and 3). Our study highlights the considerable po-

tential to add structure in unstructured citizen science initiatives, by promoting an adaptive sampling protocol focused on a specific aim. Such adaptive sampling can increase the value of citizen science data for biodiversity research and open up new avenues for citizen science project design.

Our results contribute to the growing body of literature highlighting the potential of citizen science participants to increase the value of their sampling (Xue et al. 2016, Tiago et al. 2017b, Kays et al. 2021). For example, eBird participants were incentivized to sample undersampled areas in a game called avicaching (Xue et al. 2016). Similarly, Kays and colleagues (2021) used a "plan, encourage, supplement" approach, to improve spatial coverage of camera traps in North Carolina. However, our approach differed because sampling priority was informed by a statistical outcome (i.e., species richness estimation), used to inform citizen science participants. A similar project in the United Kingdom called DE-CIDE (https://decide.ceh.ac.uk) encourages sampling aimed at improving species distribution models. These results of the potential of behavioral change conform to other studies of behavioral change in different fields such as education (Hardy et al. 2011), sustainability (Reeves et al. 2012), and health (Barankowski et al.
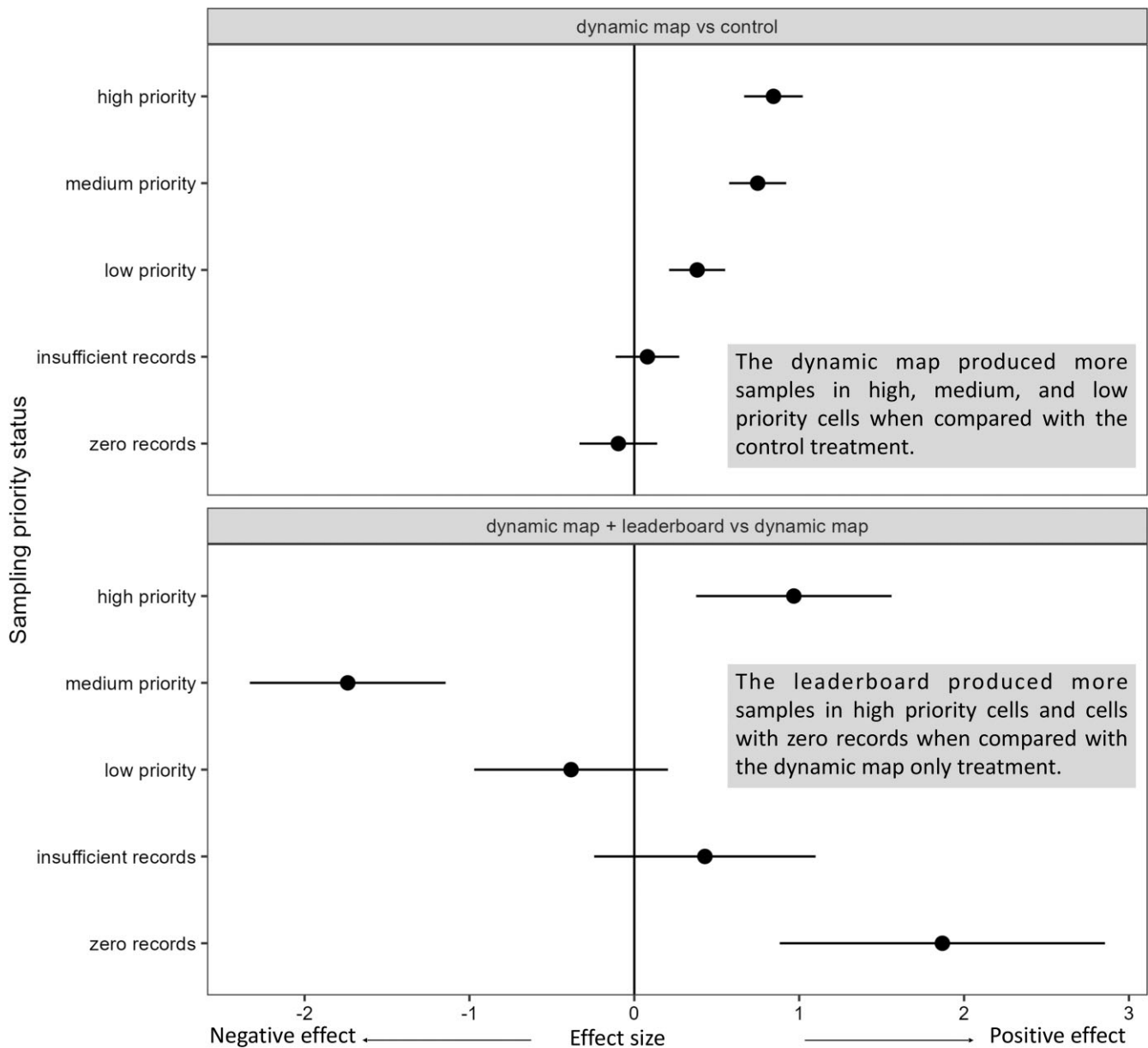
Figure 3. Model results of our mixed effect model testing two questions, stratified for each sampling priority status. Top, the effect of the dynamic map versus the control study regions with positive values indicating that the participants presented with the dynamic map increased their sampling in the given priority status relative to the control regions. Bottom, the difference between the dynamic map and the dynamic map with a leaderboard study regions, with positive values indicating the leaderboard promoted increased sampling relative to the dynamic map only treatment. Each point represents the effect size and the black line represents the 95% confidence interval. When the 95% confidence interval does not overlap zero, this can be interpreted as a significant effect of the experiment.

2008). However, to our knowledge, we are the first to experimentally test the efficacy of behavioral nudges in a citizen science initiative.

Conservation knowledge is a major motivation for many citizen science participants (Tiago et al. 2017b, Maund et al. 2020, Anđelković et al. 2022), and therefore, we speculate that the participants' willingness to adapt different sampling strategies probably reflects motivations to contribute to conservation. But there is a wide diversity of citizen science participants (August et al. 2020, Di Cecco et al. 2021, West et al. 2021), with different motivations for participating in citizen science projects (Maund et al. 2020) that can vary among different socioeconomic groups (West et al. 2021) and even among citizen science initiatives (Agnello et al. 2022).

We found mixed support of the influence of the presence of a leaderboard, focused on the most valuable observations, as opposed to the number of observations or most species observed, often displayed in citizen science projects. The leaderboard tended to increase sampling of high priority cells, but this was reversed for medium priority cells. And the leaderboard also tended to increase sampling of zero record cells as these were awarded the most points. However, zero records and insufficient records cells made up a relatively small percentage of cells, and so a small number of samples could result in a large statistical effect. Because the point value would be easier for medium (2 points) and high priority cells (3 points) it is also possible participants were maximizing the number of points on the leaderboard

appropriately. Directly asking the participants what they were maximizing for and the extent to which they used the leaderboard nudge will be important future work to understand how citizen science participants contribute to adaptive sampling protocols. On the basis of our familiarity with the region, we speculate that some of these zero records or insufficient records areas were probably more difficult to access and had less suitable habitat for frogs than the medium and high priority cells. FrogID had been running for 3 years (Rowley et al. 2019) before our study and, because most cells had already been sampled, the very small percentage of cells that were not heavily targeted despite our nudges potentially indicates particular attributes that restrict sampling. And this is why we labelled these differently than high, medium, or low priority cells. Had such an adaptive sampling protocol been initiated at the start of a citizen science project, we suspect that there would have been a stronger effect of sampling in cells with zero or insufficient records for both experimental tests instead of just the leaderboard test. Nevertheless, identifying regions that are either inaccessible or unsuitable for target species for citizen science participants can guide future professional surveys (Tulloch and Szabo 2012, Tulloch et al. 2013, Kays et al. 2021). Ultimately, we found uncertainty about the leaderboard's influence, and the potential nuance of using a leaderboard to encourage adaptive sampling. This finding may reflect people's intrinsic motivation for contributing to a citizen science project, as opposed to external validation of their contributions (Feng et al. 2018). In other words, these results, although they were mixed, provide support to the notion that self-efficacy (Feng et al. 2018) or personal reasons (West et al. 2021) are less motivational in behavioral nudges than intrinsic motivations. This finding supports other work that surveyed environmental citizen scientists and that showed that participant motivations were dominated by concern for the environment and aligned with project goals (Larson et al. 2020, West et al. 2021). However, our study did not investigate how the results (e.g., influence of the leaderboard) changed among demographic groups. Motivations for participation in citizen science can differ among demographic and socioeconomic groups (Pateman et al. 2021, West et al. 2021), and this remains an important avenue of further research in fully quantifying the potential of adaptive sampling protocols.

An additional avenue for future research that our study highlights is how we conceptualize spatial sampling bias in citizen science data. On analyzing our data, we realized that a simple measure of spatial bias (e.g., Moran's I) would not be satisfactory, because our design implicitly introduced spatial bias, given that species richness is not uniformly. Therefore, for our objective of maximizing understanding of species richness, a simple reduction in spatial bias would not necessarily be informative. Because this was the case, we developed an approach of using a null model (see figure 4) that we believe may represent a potential future way to assess changes in adaptive sampling by quantifying the change in spatial clustering of observations.

Citizen science participants' motivations, activity levels, and skill all vary among different projects and among taxa (Bowler et al. 2022). Although our study was focused on frogs (the FrogID citizen science project), our results are widely applicable, given the intrinsic interests of many participants who contribute to citizen science (Maund et al. 2020). For example, although FrogID collects data at points, we aggregated these data into grids, allowing for some generalizability to other citizen science projects that collect data through transects or area-based searches (e.g., breeding bird surveys). During our 1-year study, there were COVID-19 lockdowns, which were known to have altered citizen science

reporting (e.g., Rose et al. 2020, Sánchez-Clavijo et al. 2021), potentially limiting travel to distant high priority or medium priority cells. Therefore, our results may be a conservative estimate of the potential for adaptive sampling in citizen science projects. We also note that not all sampling priority statuses from all study regions followed the general trend; for example, the Central Coast study region had only 8% of samples from high priority cells. This was because of one superuser who submitted numerous records from what was generally a medium priority cell throughout the study. If this superuser was removed from the analysis the results were qualitatively (supplemental figure S10) and quantitatively (supplemental figure S11) similar, but the overemphasis of medium priority sampling was indeed minimized. Such biases can be introduced by heavy users, which highlights the difficulties in working with observation data in an experimental setting. Nevertheless, our statistical test and empirical interpretation shows a consistent trend when the experimental study regions are compared with the control regions.

Biodiversity monitoring with citizen science data is not only about tracking diversity or species richness, as was the focus of our study. Citizen science data are increasingly broadly used, including species distribution modeling (Milanesi et al. 2020), discovery or rediscovery of rare and new species (e.g., Vendetti et al. 2018, Richart et al. 2019), monitoring alien populations (Dart et al. 2022), and tracking population abundance changes (Horns et al. 2018, Gorta et al. 2019). Although we focused on estimating species richness, our results suggest that our approach could also apply to improving information for these other applications. Future adaptive sampling schemes could focus on understanding and capitalizing on the different motivations of citizen science participants to improve the effectiveness of citizen science initiatives. This could involve a focus on multiple types of objectives (e.g., sampling undersampled regions, estimating population change, discovery of new or missing species), each supported by a different map of sampling priority, because these will inherently have different optimal sampling requirements in space and time (Callaghan et al. 2019b), depending on the chosen goal or objective. The participants could then opt in to receive updates about their particular goal or objective. Such an approach might be particularly attractive to the superusers (supplemental figure S8; Wood et al. 2011, Rowley et al. 2019), the relatively small percentage of users who contribute the majority of data. And whether and to what extent different subsets of users adopt different nudges remains an important avenue for future research.

Data to make informed decisions in an applied management sense (e.g., ecological restoration and biodiversity conservation) are increasingly important, and citizen science continues to increase and provide the potential for such data (Peters et al. 2015, Fraisl et al. 2020, Bonney et al. 2021). Our research has clear implications for applied management at local to regional scales. Not only are these data valuable generally, but information can be maximized by having a clear management objective and either collecting data to make the most informed decision or tracking the effectiveness of local-scale management decisions. For example, practitioners who implement restoration aimed at increasing the local species pool could encourage sampling that is dedicated to sampling the species diversity in the region. Furthermore, our work helps to set the scene for future research that can experimentally test and validate the robustness of incorporating adaptive sampling into citizen science design and implementation. As an example, future work could look at how annotation or the lack of it on maps can influence the likelihood of the participants to adopt behavioral nudges. Similarly, different objectives
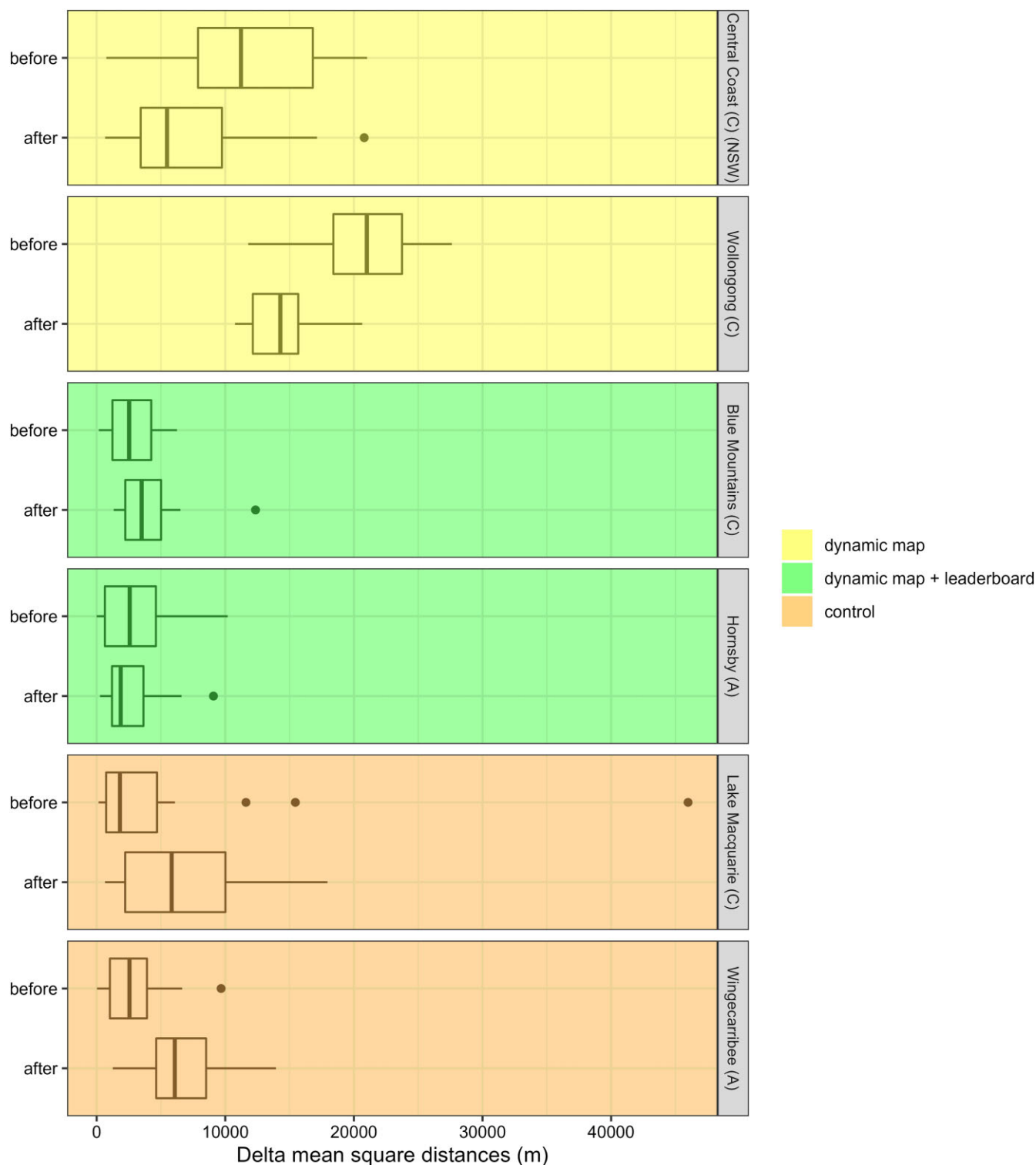
Figure 4. The difference in spatial bias measures as the mean square distance from all submissions in a study region compared with a null model of what spatial bias should look like had the dynamic maps been sampled. We generated pseudonull models for before by generating a dynamic map on the basis of our adaptive sampling protocol described in the text.

(i.e., species distribution modelling, biodiversity change through time) could be experimentally tested to see whether there are different responses to behavioral nudges among different scientific objectives.

People form the foundation of citizen science initiatives, and understanding how they engage with nature and how their moti-

vations can be harnessed for improved biodiversity conservation remains an important aspect of future research in the citizen science field (Maund et al. 2020). We showed that citizen science participants are willing to change the focus of their observations with behavioral nudges, to help achieve a specific aim. The power of adaptive sampling lies in creating a data-based feedback loop with

Figure 5. Examples of communications we received from the FrogID participants illustrating the potential and interest people have in helping contribute more meaningful citizen science sampling.

a specific aim or question and encouraging dynamic sampling to understand that aim. Citizen science projects can be multifaceted with evolving scientific objectives, goals, or questions, allowing the participants to join the scientists on that journey. In addition to our quantitative evidence, we have found much qualitative evidence illustrating the excitement and willingness of citizen science participants to be as helpful as possible for scientific goals of the FrogID project (see figure 5). Leveraging this excitement of citizen science participants, in a dynamic and adaptive framework, is a meaningful and impactful way to increasingly advance our collective knowledge of biodiversity.

Acknowledgments

Supplemental data

Supplemental data are available at *BIOSCI* online.

References cited

Agnello G, Vercammen A, Knight AT. 2022. Understanding citizen scientists' willingness to invest in, and advocate for, conservation. *Biological Conservation* 265: 109422.

Anđelković AA, Handley LL, Marchante E, Adriaens T, Brown PMJ, Tricarico E, Verbrugge LNH. 2022. A review of volunteers' motivations to monitor and control invasive alien species. *NeoBiota* 73: 153–175.

August T, Fox R, Roy DB, Pocock MJ. 2020. Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. *Scientific Reports* 10: 1–12.

Baranowski T, Buday R, Thompson DI, Baranowski J. 2008. Video games and stories for health-related behavior change. *American Journal of Preventive Medicine* 34: 74–82.

Bayraktarov E, Ehmke G, O'Connor J, Burns EL, Nguyen HA, McRae L, Possingham HP, Lindenmayer DB. 2019. Do big unstructured biodiversity data mean more knowledge? *Frontiers in Ecology and Evolution* 6: 239.

Billaud O, Vermeersch R, Porcher E. 2021. Citizen science involving farmers as a means to document temporal trends in farmland biodiversity and relate them to agricultural practices. *Journal of Applied Ecology* 58: 261–273.

Bird TJ, et al. 2014. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation* 173: 144–154.

Boakes EH, McGowan PJ, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM. 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLOS Biology* 8: e1000385.

Bonney R, Byrd J, Carmichael JT, Cunningham L, Oremland L, Shirk J, Von Harten A 2021. Sea Change: Using citizen science to inform fisheries management. *BioScience* 71: 519–530.

Bowler D, et al. 2022. Decision-making of citizen scientists when recording species observations. *Scientific Reports* 12: 11069.

Bowser A, Hansen D, He Y, Boston C, Reid M, Gunnell L, Preece J. 2013. Using gamification to inspire new citizen science volunteers. Pages 18–25 in Lennart E, Nacke K Harrigan NR, eds. *Proceedings of the First International Conference on Gameful Design, Research, and Applications*. Association for Computing Machinery.

Callaghan CT, Poore AG, Major RE, Rowley JJ, Cornwell WK. 2019a. Optimizing future biodiversity sampling by citizen scientists. *Proceedings of the Royal Society B* 286: 20191487.

Callaghan CT, Rowley JJ, Cornwell WK, Poore AG, Major RE. 2019b. Improving big citizen science data: Moving beyond haphazard sampling. *PLOS Biology* 17: e3000357.

Callaghan CT, Roberts JD, Poore AG, Alford RA, Cogger H, Rowley JJ. 2020. Citizen science data accurately predicts expert-derived species richness at a continental scale when sampling thresholds are met. *Biodiversity and Conservation* 29: 1323–1337.

Chandler M, et al. 2017. Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation* 213: 280–294.

Chao A, Gotelli NJ, Hsieh T, Sander EL, Ma K, Colwell RK, Ellison AM. 2014. Rarefaction and extrapolation with hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs* 84: 45–67.

Courter JR, Johnson RJ, Stuyck CM, Lang BA, Kaiser EW. 2013. Weekend bias in citizen science data reporting: Implications for phenology studies. *International Journal of Biometeorology* 57: 715–720.

Dart K, Latty T, Greenville A 2022. Citizen science reveals current distribution, predicted habitat suitability and resource requirements of the introduced African carder bee *Pseudoanthidium (Immanthidium) repetitum* in Australia. *Biological Invasions* 24: 1827–1838.

Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH. 2021. Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *BioScience* 71: 1179–1188.

Eveleigh A, Jennett C, Blandford A, Brohan P, Cox AL. 2014. Designing for dabblers and deterring drop-outs in citizen science. Pages 2985–2994 in Jones M, Palanque P, Schmidt A Grossman T, eds. *CHI'14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.

Feng Y, Ye HJ, Yu Y, Yang C, Cui T. 2018. Gamification artifacts and crowdsourcing participation: Examining the mediating role of intrinsic motivations. *Computers in Human Behavior* 81: 124–136.

Fontaine C, Fontaine B, Prévot A-C. 2021. Do amateurs and citizen science fill the gaps left by scientists? *Current Opinion in Insect Science* 46: 83–87.

Forister M, Halsch C, Nice C, Fordyce J, Dilts T, Oliver J, Prudic K, Shapiro A, Wilson J, Glassberg J. 2021. Fewer butterflies seen by community scientists across the warming and drying landscapes of the American West. *Science* 371: 1042–1045.

Fraisl D, et al. 2020. Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science* 15: 1735–1751.

Geldmann J, Heilmann-Clausen J, Holm TE, Levinsky I, Markussen B, Olsen K, Rahbek C, Tøttrup AP. 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions* 22: 1139–1149.

Gorta SB, et al. 2019. Pelagic citizen science data reveal declines of seabirds off south-eastern Australia. *Biological Conservation* 235: 226–235.

Haklay M. 2013. Citizen science and volunteered geographic information: Overview and typology of participation. Pages 105–122 in Sui D, Elwood S Goodchild M, eds. *Crowdsourcing Geographic Knowledge*. Springer.

Hardy JL, Drescher D, Sarkar K, Scanlon M. 2011. Enhancing visual attention and working memory with a Web-based cognitive training program. *Mensa Research Journal* 42: 13–20.

Horns JJ, Adler FR, Şekercioğlu ÇH. 2018. Using opportunistic citizen science data to estimate avian population trends. *Biological Conservation* 221: 151–159.

Hsieh T, Ma K, Chao A 2016. iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* 7: 1451–1456.

Isaac NJ, Strien AJv, August TA, Zeeuw MPd, Roy DB. 2014. Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution* 5: 1052–1060.

Johnston A, Moran N, Musgrove A, Fink D, Baillie SR. 2020. Estimating species distributions from spatially biased citizen science data. *Ecological Modelling* 422: 108927.

Kays R, Lasky M, Parsons AW, Pease B, Pacifici K 2021. Evaluation of the spatial biases and sample size of a statewide citizen science project. *Citizen Science: Theory and Practice* 6: 34.

Kirchhoff C, Callaghan CT, Keith DA, Indiarto D, Taseski G, Ooi MK, Le Breton TD, Mesaglio T, Kingsford RT, Cornwell WK. 2021. Rapidly mapping fire effects on biodiversity at a large-scale using citizen science. *Science of the Total Environment* 755: 142348.

Larson LR, Cooper CB, Futch S, Singh D, Shipley NJ, Dale K, LeBaron GS, Takekawa JY. 2020. The diverse motivations of citizen scientists: Does conservation emphasis grow as volunteer participation progresses? *Biological Conservation* 242: 108428.

Manzano-León A, Camacho-Lazarraga P, Guerrero MA, Guerrero-Puerta L, Aguilar-Parra JM, Trigueros R, Alias A. 2021. Between level up and game over: A systematic literature review of gamification in education. *Sustainability* 13: 2247.

Maund PR, Irvine KN, Lawson B, Steadman J, Risely K, Cunningham AA, Davies ZG. 2020. What motivates the masses: Understanding why people contribute to conservation citizen science projects. *Biological Conservation* 246: 108587.

Milanesi P, Mori E, Menchetti M. 2020. Observer-oriented approach improves species distribution models from citizen science data. *Ecology and Evolution* 10: 12104–12114.

Morford ZH, Witts BN, Killingsworth KJ, Alavosius MP. 2014. Gamification: The intersection between behavior analysis and game design technologies. *Behavior Analyst* 37: 25–40.

Pateman R, Dyke A, West S. 2021. The diversity of participants in environmental citizen science. *Citizen Science: Theory and Practice* 6: 9.

Peters MA, Eames C, Hamilton D. 2015. The use and value of citizen science data in New Zealand. *Journal of the Royal Society of New Zealand* 45: 151–160.

Pocock MJ, Roy HE, Preston CD, Roy DB. 2015. The Biological Records Centre: A pioneer of citizen science. *Biological Journal of the Linnean Society* 115: 475–493.

Pocock MJ, Tweddle JC, Savage J, Robinson LD, Roy HE. 2017. The diversity and evolution of ecological and environmental citizen science. *PLOS ONE* 12: e0172579.

Reeves B, Cummings JJ, Scarborough JK, Flora J, Anderson D. 2012. Leveraging the engagement of games to change energy behavior. Pages 354–358 in Fox G Smari WW, eds. *2012 International*

Conference on Collaboration Technologies and Systems (CTS). Institute of Electrical and Electronics Engineers.

Richart CH, Chichester LF, Boyer B, Pearce TA. 2019. Rediscovery of the southern California endemic American keeled slug *Anadenu-luscockerelli* (Hemphill, 1890) after a 68-year hiatus. *Journal of Natural History* 53: 1515–1531.

Rose S, Suri J, Brooks M, Ryan PG. 2020. COVID-19 and citizen science: Lessons learned from southern Africa. *Journal of African Ornithology* 91: 188–191.

Rowley JJ, Callaghan CT. 2020. The FrogID dataset: Expert-validated occurrence records of Australia's frogs collected by citizen scientists. *ZooKeys* 912: 139.

Rowley JJ, Callaghan CT, Cutajar T, Portway C, Potter K, Mahony S, Trembath DF, Flemons P, Woods A. 2019. FrogID: Citizen scientists provide validated biodiversity data on frogs of Australia. *Herpetological Conservation and Biology* 14: 155–170.

Sánchez-Clavijo LM, et al. 2021. Differential reporting of biodiversity in two citizen science platforms during COVID-19 lockdown in Colombia. *Biological Conservation* 256: 109077.

Schuster R, et al. 2019. Optimizing the conservation of migratory species over their full annual cycle. *Nature Communications* 10: 1–8.

Takahashi A, Kumagai Y, Aoki H, Tamura R, Oba F. 2022. Adaptive sampling methods via machine learning for materials screening. *Science and Technology of Advanced Materials* 2: 55–66.

Thaler RH, Sunstein CR. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin.

Tiago P, Ceia-Hasse A, Marques TA, Capinha C, Pereira HM. 2017a. Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Scientific Reports* 7: 1–9.

Tiago P, Gouveia MJ, Capinha C, Santos-Reis M, Pereira HM. 2017b. The influence of motivational factors on the frequency of participation in citizen science activities. *Nature Conservation* 18: 61–78.

Tulloch AI, Szabo JK. 2012. A behavioural ecology approach to understand volunteer surveying for citizen science datasets. *Emu-Austral Ornithology* 112: 313–325.

Tulloch AI, Mustin K, Possingham HP, Szabo JK, Wilson KA. 2013. To boldly go where no volunteer has gone before: Predicting volunteer activity to prioritize surveys at the landscape scale. *Diversity and Distributions* 19: 465–480.

Vendetti JE, Lee C, LaFollette P. 2018. Five new records of introduced terrestrial gastropods in Southern California discovered by citizen science. *American Malacological Bulletin* 36: 232–247.

Welvaert M, Caley P. 2016. Citizen surveillance for environmental monitoring: Combining the efforts of citizen science and crowdsourcing in a quantitative data framework. *SpringerPlus* 5: 1–14.

West S, Dyke A, Pateman R. 2021. Variations in the motivations of environmental citizen scientists. *Citizen Science: Theory and Practice* 6: 14.

Wood C, Sullivan B, Iliff M, Fink D, Kelling S. 2011. eBird: Engaging birders in science and conservation. *PLOS Biology* 9: e1001220.

Xue Y, Davies I, Fink D, Wood C, Gomes CP. 2016. Avicaching: A two stage game for bias reduction in citizen science. Pages 776–785 in Thangarajah J, Tuyls K, Jonker C Marsella S, eds. *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems.

Yocoz NG, Nichols JD, Boulinier T. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16: 446–453.

Zeng Y, Xiang K. 2017. Adaptive sampling for urban air quality through participatory sensing. *Sensors* 17: 2531.