Title: Assessing veracity of web-based citizen science projects: An audit study of eBird

Authors: Henry M. Streby[1], Anonymous[2], Anonymous[3], and Anonymous[4]

Author affiliations: [1]Department of Environmental Sciences, University of Toledo, Toledo, Ohio, USA

[2-4] Undisclosed

Acknowledgements: We thank eleven anonymous colleagues for insightful feedback on the manuscript. We also thank three anonymous ex-eBird associates for sharing insights into the eBird data quality control process. This project was completed without funding. Three of us will remain publicly anonymous out of concern for personal and professional retaliation. We thank the University of Toledo chapter of the American Association of University Professors for supporting HMS, and the many leading experts in all areas relevant to our study for providing testimony unanimously confirming the legitimacy and importance of our work.

Ethics Statement: This study follows standard web-based auditing methods that have been reviewed and confirmed to be consistent with the practices of the relevant research community. Similar ethics language is included in the manuscript, but this statement is added here at the request of the Editors so it is up-front for the reader. Web-based audit studies involve the submission, over the Internet, of fake/fictitious information to a platform in order to test how the platform responds to the experimental treatment of the fictitious information (Sandvig 2014). Audit studies, by definition, involve deception (i.e., the audited entity cannot be aware of the experimental treatment) to ensure the data (i.e., the entity's responses) are representative and independent of internal influence. Notably, auditing researchers are not bound by website terms-of-use that prohibit users from submitting false information, and such action is not fraud or hacking (Sandvig et al. v Barr 2020). This study did not include human research subjects. The study of demographic variation in fictitious people does not require Institutional Review Board (IRB) oversight. Further, manipulating an environment, and or interacting with humans, does not make them human research subjects under the Common Rule (HHS 45 CFR 46, Subpart A) without also obtaining information *about* specific individuals or collecting biospecimens from them. For the current study, "The experimental plan does **not** fall within the purview of requiring review by an Institutional Review Board as per the Common Rule" (E. Bankert, Director of IRB, Dartmouth College, expert testimony to UT Admin; emphasis her own). Despite confirmation of this fact before and after our study from UToledo IRB analysts, unanimous testimony confirming no need for IRB oversight from leading experts in the Common Rule (E. Bankert) and audit study design (C. Sandvig, expert testimony to UToledo admin), and a subsequent finding by a UT Investigation Committee that the IRB "made a mistake," the UToledo IRB prohibited the use of 0.5% of our data following undisclosed communications between UToledo admin and others outside UT. IRBs are federally mandated entities, and their decisions cannot be overturned by anyone but themselves. We are therefore prohibited from reporting results regarding bias based on region, sex, or race. These important issues should be investigated in future research.

ABSTRACT

1. Uncertainty in data accuracy and lack of diversity among participants are two common concerns in citizen science projects. But what if the methods used to simultaneously maximize data quality and user participation lead to unbalanced user experiences and biased data? While many studies focus on advancing techniques for analyzing large and noisy datasets, methods are needed to test the veracity of the systems that produce those data.

2. Here we present a web-based audit study, a standard and increasingly common method in computer and social sciences but novel to ecology, to test the veracity of methods used in web-based citizen science data collection and curation. We tested for several types of bias in eBird's data quality control process, with the intent to provide recommendations to improve representation and scientific veracity in the world's largest citizen science project. Following a basic design for auditing web-based platforms, we systematically submitted 19,491 fictitious bird observations from 196 simulated eBirders who varied by region, sex, and race, and we assessed how those variables were associated with eBird's real acceptance or rejection of the fictitious observations.

3. We found near-total confirmation bias, in that eBird accepts 100% of false observations that fit eBird's *a priori* expectations for species' spatio-temporal ranges, while accepting 0 – 29% of true observations that do not fit eBird's expectations. Hence, the eBird data quality control process does not validate data, but rather it produces illusory quality through forced conformity. We show how this problem contributes to a larger, self-perpetuating cycle of bias inherent to the design and function of eBird. We cannot report results from geographic or demographic analyses (see SI).

4. Regarding species' spatio-temporal ranges, and changes thereto, we conclude that the eBird database almost entirely reflects eBird's predetermined expectations, and changes that eBird makes to those expectations over time. This irreparably detaches the eBird dataset from ecological reality because no statistical technique can reconstitute patterns that are perpetually and permanently removed and replaced during data collection and curation. eBird is an undeniably magnificent, user-friendly birdwatching application, but as citizen science, its design obviates the citizen and feigns science. The ecological and evolutionary sciences could benefit from similar independent investigations of other web-based research projects.

Key-words: algorithm auditing, big data, confirmation bias, discrimination, eBird, ornithology

1 | INTRODUCTION

Citizen science (aka, community science), or the involvement of volunteers in empirical research, has enabled the collection of enormous volumes of data by reducing financial, logistical, or physical constraints inherent to traditional methods (Dickinson et al. 2010). In ecology, some of the largest and longest-running citizen science projects include the North American Breeding Bird Survey, National Audubon Society's Christmas Bird Count, North American Amphibian Monitoring Program, the leafing and flowering times of US lilacs and honeysuckles, and the Butterfly Monitoring Scheme (Kosmala et al. 2016). Data collected from these and other citizen science projects have advanced the understanding of phenological responses to climate change (Hurlbert & Liang 2012), long-term population trends (Sauer et al. 2017), descriptions of animal and plant distributions (Braschler 2009, Deguines et al. 2012), spread of invasive species (e.g., Delaney et al. 2008), and effects of anthropogenic changes to landscapes (Ward-Paige et al. 2010).

eBird (ebird.org; Cornell Lab of Ornithology [hereafter, CLO]) is a relative newcomer (2002) to the citizen science scene that has eclipsed all other biodiversity-related projects in terms of global coverage and participation rates. eBird capitalizes on the widespread popularity of birdwatching and has compiled >2 billion observations submitted by birdwatchers (hereafter, eBirders), over the Internet through eBird's free website and mobile application, into a central database (Sullivan et al. 2014, ebird.org). eBirders include nine-year olds, retired hobby birdwatchers, and everyone in between (J. Fitzpatrick, CLO, pers. comm. to University of Toledo [hereafter, UT admin]). The eBird database has been used to show shifting species distributions (Taylor et al. 2014), responses to climate change (Reese & Skagen 2017), population trends (Clark 2017, Walker & Taylor 2017), changes in migration patterns and phenology (Supp et al. 2015), and to estimate global abundances of many bird species (Callaghan et al. 2021). Other applications of eBird data include "species management, habitat protection, and informing law and policy" (ebird.org).

Citizen science data, such as those submitted by eBirders, have enormous potential scientific value, but may have limited utility for some purposes because volunteer-collected observational data can be subject to greater rates of human error and/or bias than those collected by trained and supervised employees (Foster-Smith & Evans 2003, Crall et al. 2011, Gardiner et al. 2012, Kamp et al. 2016, but see Kosmala et al. 2016). Data quality issues are most concerning when user submissions cannot be verified by the citizen science project or the researchers using the resulting dataset, as is the case with eBird. Other web-based citizen science projects such as zooniverse (zooniverse.org) and iNaturalist (inaturalist.org) receive submissions comprised primarily of photographs, which are reviewed by additional citizen scientists and researchers for accuracy before being accepted into the database, and provide a record allowing subsequent re-examination. Despite recent upgrades to eBird that allow photographs and other media to accompany eBirder submissions, most observations lack such documentation, likely due to the challenges of photographing free-living birds and the prohibitive cost of quality photography equipment for many users. In the absence of evidentiary documentation, eBird reports are eyewitness accounts, and a large body of literature documents that human eyewitnesses are often more confident than they are reliable (e.g., Rose & Beck 2014).

eBird recognizes the user accuracy issues inherent to systems that accept unverifiable anecdotes from untrained non-experts, as evidenced by its implementation of a two-tier data quality control process (Sullivan et al. 2014). Briefly, all observations submitted to eBird first go through algorithm-based filters that essentially classify each observation as expected (i.e., the species is reported within a spatial and temporal range that is consistent with eBird's *a priori* expectations) or unexpected (i.e., the species is reported outside of eBird's expected spatial and/or temporal range, by any magnitude). It is unclear if these filters take anything else into consideration, and the algorithms are not publicly available. These expectation filters were originally set by eBird associates including volunteer eBird reviewers ("regional experts;" Sullivan et al. 2014, pg. 34), and can be adjusted at any time by those reviewers. All expected observations are accepted into the database automatically. eBird reviewers can scrutinize and reject expected observations that were automatically accepted by filters, but this is rare. All observations that are classified as unexpected by an algorithm-based filter require the eBirder to confirm that they entered the intended species by adding a note about their observation during data submission, and the observation will then purportedly receive subsequent scrutiny by eBird reviewer(s). eBirders are said to be given the opportunity to explain each unexpected observation within a few days of submission (J. Fitzpatrick, CLO, pers. comm. to UT admin), but unexpected observations are commonly accepted or rejected without inquiry from eBird reviewers (the authors, pers. obs.). Accepted observations, whether expected or unexpected, enter into the eBird database and are subsequently viewable by anyone, in map and checklist form, and available to extract for scientific analysis. Rejected observations are omitted from the eBird database, not publicly viewable or extractable for analysis, and typically removed from the eBirder's account, making them permanently irretrievable (i.e., eBird keeps no backups of data lost or removed from user accounts; eBird.org). Despite the intent of the data quality control process, eBird's selective acceptance of evidence that supports *a priori* expectations while rejecting evidence that refutes those expectations, regardless of the validity of the evidence, meets the definition of confirmation bias, and might have unintended consequences for the quality and utility of the resulting dataset.

While eBird's algorithms and volunteer reviewers work diligently to maximize data quality, the CLO may also be unmatched in its efforts to maximize participation in eBird. Increasing geographic coverage and increasing the diversity of eBirders are admirable goals of the program (ebird.org), and there is much work to be done. For example, in the US, 95% of eBirders are non-Hispanic White (Rutter et al. 2021), and in metropolitan areas, eBird data are spatially biased and overrepresent wealthier neighborhoods with higher proportions of White residents

(Perkins 2020, Grade et al. 2022). Even in majority non-White regions of the world, the most active eBird contributors are typically White (Scott 2021), and eBird's design and global standardization of bird names can "contribute to the erasure of cultural diversity" and "reinforce the *status quo* of colonizer–colonized relationships" (Soares et al. 2023). There is growing recognition within the birdwatching and scientific communities that implicit biases result in inequities in awareness, opportunity, and participation based on race, sex, and other demographic parameters (e.g., Gentile 2019, Thompson 2020, Soares et al. 2023). Recognition of these biases has led to initiatives and organizations including Black Birders Week (blackafinstem.com), Bird Names for Birds (birdnamesforbirds.wordpress.com), movements to address sexism (e.g., Ortega 2021), and calls for policy changes that encourage respectful collaboration with international partners (Soares et al. 2023). But what happens if a research entity's efforts to increase data quantity and quality introduce more opportunity for geographic or demographic biases to arise? Disproportionate treatment of participants based on their region, sex, race, or other factors could undermine efforts to increase inclusiveness and more representative data, and identifying and working to correct such biases should be desired by all.

Despite discussion and debate about the pros and cons of citizen science (e.g., Crall et al. 2011, Gardiner et al. 2012, Brown & Williams 2018), the reliability of the resulting data (e.g., Foster-Smith & Evans 2003, Crall et al. 2011), and efforts to quantify the success of internal review processes for filtering erroneous data (Bonter & Cooper 2012), to our knowledge, there have been no attempts to quantify potential biases introduced by the data quality control process itself within a citizen science project. Investigating the specific types of bias for which we were concerned (i.e., confirmation bias and bias based on geographic region, sex, and race), in a web-based platform, requires specific computer science methods. We implemented an audit study (Sandvig et al. 2014), which is a standard and increasingly common method in the computer sciences and social sciences (Gaddis 2018). Web-based audits have been used in the medical/natural sciences to test, for example, the accuracy of online symptom checkers (Semigran et al. 2015), but the methods may be novel to many ecologists. Audit study designs, considered “field studies” in computer and social sciences, have been in use since at least 1940, but their application to web-based platforms has grown with increasing use of the Internet (Gaddis 2018). Briefly, web-based audit studies involve the submission, over the Internet, of fake/fictitious information to a platform in order to test how the platform responds to the experimental treatment of the fictitious information (Sandvig 2014). Audit studies, by definition, involve deception (i.e., the audited entity cannot be aware of the experimental treatment) to ensure the data (i.e., the entity’s responses) are representative and independent of internal influence. Notably, auditing researchers are not bound by website terms-of-use that prohibit users from submitting false information, and such action is not fraud or hacking (Sandvig et al. v Barr 2020). Audit studies have been used to uncover discrimination in treatment of rideshare users (Ge et al. 2016), racial bias in healthcare algorithms (Obermeyer et al. 2019), and failures of social media platforms to identify false political ads (Le Pochat et al. 2022), among many other applications.

We used standard web-based audit methods to test for several types of bias within the eBird data quality control process that might undermine efforts to increase participant diversity and could impact the scientific veracity of the database. Specifically, we tested for confirmation bias and bias based on region, sex, and race of eBirders. We systematically submitted fictitious observations representing expected and unexpected times and locations from simulated eBirder accounts for which the eBirders represented regional and demographic variation, to test if those parameters were associated with the rates at which observations were accepted into the eBird database. As in other audit studies, although our submissions were fictitious, they represented real submissions, their treatment by eBird was real and representative, and our analyses and results are based on the real final status of each observation (i.e., accepted into the eBird database or not). Our intention was to help improve the inclusiveness and scientific veracity of the world's largest citizen science project by identifying and quantifying potential forms of bias that could be reduced or accounted for by changes in eBird's approach to data collection and curation. We predicted that acceptance and rejection rates would be similar across regions but might vary with demographic identities of eBirders. In addition, we predicted that confirmation bias would be present at some rate but would be small enough to be remedied by minor adjustments in eBird's practices. We used our findings to assess the utility of eBird data for delineating species' spatio-temporal ranges and detecting changes in those ranges.

2 | METHODS

We created eBird accounts for 196 simulated eBirders, through which we submitted lists of fictitious bird observations (i.e., real species, but not real observations). Following audit research standard ethics (e.g., Sandvig et al. 2014), we designed our study to reduce, and in this case eliminate, possible influence of our intervention on anyone else interacting with eBird (see SI). We took care to make our simulated eBirders indistinguishable from typical eBirder accounts and activities, and to be similar to each other in all ways except the variables we meant to test. To make each simulated eBirder appear ordinary, we uploaded a randomly generated list of common species that they "observed" prior to joining eBird (i.e., life lists) and we regularly submitted lists (~2/mo) of fictitious observations that were expected (by eBird) species during February – May 2017. We submitted these bird lists as if they were from locations at or near where presumably real lists (i.e., lists from existing eBirders are unverifiable by us or anyone else) had been recently submitted by eBirders. We created specific differences among our simulated eBirders to test for geographic and demographic biases. For regional variation, we assigned four simulated eBirders to each continental U.S. state. For demographic variation, each of the four simulated eBirders in each state was assigned a name socially perceived as belonging to one of four groups based on sex (i.e., typically male vs. typically female) and race (White-sounding vs. Black-sounding; See SI).

Our submission of several fictitious bird lists from each simulated eBirder that contained only species expected by eBird served three purposes: 1) to establish a level of experience that was similar among all of our simulated eBirders and thus limited potential inexperience bias from eBird; 2) to ensure the proportion of expected and unexpected observations we submitted was close to the regular rate of eBird submissions (94% of submissions are expected [S. Kelling, CLO, pers. comm. to UT admin]); and 3) to attempt to obtain adequate variation in rejection rates of expected observations, which we anticipated would be uncommon, to compare among our variables of interest. Following the submissions that contained only expected species, in one of each simulated eBirder's reports during April or May 2017, we submitted another list containing expected observations with a single additional observation of an unexpected, but plausible, species (i.e., a bird slightly outside of eBird's expected spatial or temporal range for that species). We assigned these unexpected fictitious observations immediately adjacent to eBird's spatial or temporal expectations for two reasons: to eliminate the possibility of our entries impacting anyone else's research (see SI), and to ensure that the resulting acceptance rate would be a maximum (i.e., we assumed eBird's acceptance rate of unexpected observations peaks proximal to their expected range and decreases to zero when increasingly far off).

During the data submission process, the eBird filtering algorithms identified each unexpected observation and required an additional note to be entered prior to submission of the species list. To meet this requirement, we randomly selected a phrase from a response bank we prepared based on common eBirder entries we learned from ex-eBird reviewers (see SI). Following submission, all unexpected observations are purportedly scrutinized by eBird reviewer(s) (Kelling et al. 2015) who ostensibly request additional information within few days to corroborate observations (J. Fitzpatrick, CLO, pers comm. to UT admin). We monitored the status of all fictitious observations and responded to inquiries regarding unexpected observations in a pre-determined, standardized fashion (see SI). In October 2017, ~5 months after submission of the last list for each simulated eBirder, we recorded the status of each fictitious observation as accepted (i.e., present in eBird database) or rejected (i.e., absent from eBird database).

We anticipated that eBird's data quality control process would be rapid, as advertised, but 34% of our unexpected observations received no apparent scrutiny in the months following submission. Therefore, we left all observations, expected and unexpected, in the simulated eBirder accounts for an extended duration because deleting those that were already accepted would be atypical behavior for eBirders and could have influenced eBird's treatment of remaining unexpected observations. After waiting an additional year (i.e., October 2018), we observed no new inquiries to our simulated eBirders, and no changes to the status of any observations. We subsequently checked the status of each observation prior to removing all of them permanently from eBird (January 2020), and we again observed no additional changes in their status prior to removal.

3 | RESULTS

We submitted 1,188 checklists containing 19,491 fictitious observations to eBird, of which 19,296 were expected and 195 were unexpected based on eBird's spatial and temporal filters for each species. Excluded from this total was one checklist from one simulated eBirder in which we made an eBird entry error (i.e., wrong state). We later excluded 103 additional observations (1 expected and 102 unexpected) to be compliant (see SI), leaving 19,388 simulated observations (19,295 expected and 93 unexpected) for which we had data in the form of final status of accepted or rejected. The excluded data had constituted ~0.5% of our original dataset, and their absence did not influence results regarding confirmation bias. However, the exclusion of those data resulted in inadequate sample sizes to present analyses and results regarding bias in eBird based on region, race, and sex. Future studies should consider similar questions.

All 19,295 (100%) expected observations were accepted into the eBird database despite none of them being true. We therefore estimate the probability of a false observation being accepted into the eBird database to be 1.00, as long as it fits eBird's expectations for the species' spatio-temporal range. Considering 94% of eBird entries are expected (S. Kelling, CLO, pers. comm. to UT admin), this result indicates that at least 94% of the entire eBird dataset was accepted into the database with zero scrutiny regardless of observation validity. Of our 93 narrowly unexpected observations, 27 (29%) were accepted into the eBird database within 2 months of submission, presumably by eBird reviewers, without inquiry to our simulated eBirders, and 66 (71%) were either rejected with no inquiry, or rejected by default because no apparent action was taken on them by eBird for >2.5 years. These bird observations were fictitious, but they were each highly plausible and our results reflect the rate at which true, but narrowly unexpected, observations are accepted. Assuming that unexpected observations are rejected at increasing rates as observations become increasingly distant from eBird's spatio-temporal expectations, we estimate the probability of true, but unexpected observations being accepted into the eBird database ranges from 0.00 to 0.29. It is possible that more unexpected observations would have been accepted at later dates, but we believe that allowing 2.5 years before we removed them from the simulated eBirder accounts was long enough to conclude that the process of scrutinizing unexpected observations is detrimentally slow. We cannot report results regarding bias based on region, sex, or race (see SI).

4 | DISCUSSION

4.1 | The need for independent assessment

Here we introduced web-based audit methods to the ecological sciences. As ecologists address issues at global scales with data increasingly gathered via public, web-based approaches (e.g., social media surveys, citizen science projects, crowdsourced weather data), much attention has been paid to improving statistical methods to handle large and noisy datasets (e.g., Kelling et al. 2015, Panchariya et al. 2015, Altwegg & Nichols 2019). However, such analytical advances tend to rely on the assumption that unwieldy sample sizes and data noisiness are the only problems. For example, Kelling et al. (2015, abstract) assumed "eBird's data submission design ensures that all data meet high standards of completeness and accuracy" before introducing their big data approach to analysis. The falsity of that statement from eBird associates demonstrates the necessity for methods to independently assess the veracity of the data-collection and curation mechanisms themselves because "You can't fix by analysis what you bungled by design" (Light et al. 1990, page v). Specifically, while random error can be addressed through increased sample size, no statistical technique can recreate patterns that were systematically removed and replaced during data collection. Self-evaluation of organizational performance can be biased, and internal criticism can be deterred, as evaluators can have their objectivity compromised by the policies and values of leadership (Cummings et al. 1988, Kassing 2008). Therefore, it is incumbent upon citizen science programs that value the veracity of their databases to welcome objective third-party assessment. Applied appropriately, audit studies provide excellent tools for independently assessing web-based data collection methods and the veracity of the resulting datasets upon which research and policy rely.

4.2 | Quality control, or simply control?

Our audit study demonstrates that the eBird data quality control process imposes near-total confirmation bias by accepting 100% of observations it expects to be true, and rejecting 71–100% of observations it does not expect to be true, regardless of the validity of any of the observations. We did not expect such a strong discrepancy. It is nonsensical to assume that statistical analysis can detect meaningful patterns through the noise of the unscrutinized anecdotes that fit eBird's expectations (i.e., 94% of eBird dataset), without also assuming the same could be done for the other 6%, and thus that eBird reviewers are charged with performing a painstaking but Sisyphean task only to provide a façade of quality control. It is also nonsensical for a program to conclude that its users are wrong >70% of the time when their contributions contradict expectations, without assuming there is a similarly massive error rate when those same users' contributions match expectations. But eBird does just that, and they take it to the extreme by assuming that eBirder "expertise" is reliably measured by the proportion of expected species an eBirder reports (Kelling et al. 2015, Johnston et al. 2018). Even when eBirders submit expected species with photographs proving their own observations to be false – up to 71.4% of photographs in some circumstances (Gorleri & Areta 2022) – those false observations are accepted into the eBird database with no scrutiny and those eBirders gain credibility within eBird for agreeing with expectations (Figure S1). The logical conclusion is that eBird's ideal state is one peopled entirely by yes-sir-eBirders who submit only information that proves eBird right. But that could be accomplished more easily by simulating data within the bounds of eBird's expectations, and it begs the question of whether eBirders serve only to justify labeling eBird as citizen science and/or for selling advertising space.

The presumption that eBird's expectations are the standard of truth, and that the relative value of contributions can be measured by rate of conformity to that standard, means that eBird functions not as a scientific data collection mechanism but rather as a self-fulfilling prophecy. The true error rate of eBirder anecdotal observations, whether expected or not, is unknowable with the current quality control process. But paradoxically to their acceptance of 100% of expected observations, eBird's rejection of >70% of unexpected observations indicates that eBird has low confidence in eBirders as eyewitnesses of free-living birds. Such low confidence may seem harsh, but coincidentally, a similar percentage (75%) of wrongful convictions overturned by DNA evidence were originally based on erroneous eyewitness testimony (innocenceproject.org).

Much like a criminal justice system susceptible to selective acceptance of unreliable eyewitness accounts, the eBird data quality control process does not validate user observations, but instead it imposes predetermined outcomes under the pretense that authority begets authenticity.

4.3 | A snowball of bias accumulation

The data quality control process is only one of several eBird features that introduce bias into the database (Figure 1). When eBirders plan birdwatching activities, a convenient starting point is to reference eBird to see when and where other eBirders have reported species of interest. This real-time ability for eBirders to inform their activities is a neat birdwatching tool, but it inherently influences when and where they travel and what species they expect to see, thereby reducing the independence of samples and encouraging the perpetuation of erroneous reports, all of which go unscrutinized. Next, when eBirders enter their observations into eBird, they are presented with a checklist of species eBird expects to be present at the relevant time and place, which relegates unexpected species to write-in status. This can cause eBirders to question true observations and defer to eBird's expectations because the CLO is a revered authority (i.e., prestige bias [Troy & Jones 2022]). Indeed, rather than provide additional details on their submissions, some eBirders reduce the number of birds reported because they have been "trained" by reviewers as to what species are "supposed to be there" (J. Fitzpatrick, pers. comm. to UT admin). This operant conditioning strips eBirders of their autonomy and is the antithesis of sound study design, by demanding rather than limiting, biased data collection. Specifically, eBird induces social desirability response bias in which subjects give responses they believe the investigator wants to hear to avoid the embarrassment of reporting truth (Bernardi and Nash 2023). The next, and likely strongest layer of bias, is introduced by the eBird data quality control process, as we quantified in this study. An additional issue we did not quantify here is that eBirders are sometimes persuaded by reviewers to change unexpected observations to species that eBird expects (the authors, pers. observ., Figure S1), leading to a realized acceptance rate of >1.0 for expected species. The bias of the quality control process is then magnified by the discrepancy in timing of acceptance between expected and unexpected observations (Figure 1).

Yet more layers of bias are introduced to the eBird database every time eBird adjusts an expectation filter and eBird reviewers use the "Filter Rerun" function (ex-eBird reviewer, pers comm. UT Admin). When a filter is adjusted such that a species becomes expected in an area where it previously was not expected, historically submitted, true observations that had been rejected as unexpected cannot be re-evaluated or recovered because eBird does not keep backups of removed observations; yet all new observations of that species, whether true or false, become expected and are therefore accepted without scrutiny. Thus, a species could have long occupied an area, but subsequent analysis of eBird data from before and after a filter adjustment, would erroneously suggest the occupancy is new (i.e., false range expansion). Additionally, when a filter is adjusted such that a species becomes no longer expected in an area where it was previously expected, all historical accepted observations in that area are newly marked as unexpected, and therefore are effectively rejected, unless and until an eBird reviewer re-accepts them into the database. This induces a kind of retroactive con*form*ation bias by adjusting historical data based on current expectations. Considering the low rate at which unexpected observations are accepted in general, and the slowness with which unexpected observations are evaluated, analyses of eBird data following such a filter adjustment could erroneously conclude that a species never occupied that area, rather than having moved from it (i.e., missed range contraction). In this way, eBird is yet further removed from scientific objectivity, and more closely resembles an Orwellian Ministry of Truth, wherein truth is defined by the current opinion of a singular authority and history is perpetually rewritten to convenience current doctrine with no record maintained that any other truth ever existed.

Lastly, the bias feedback loop is perpetuated when eBirders use the resulting database, presented in a user-friendly map-based interface, to inform their next birdwatching activity. Concerningly, the cycle is accelerated when eBird data are used to inform eBird's expectations (J. Fitzpatrick, pers. comm. to UT admin) and therefore which observations will be accepted. Additionally concerning, the CLO plans to, or perhaps already does, use artificial intelligence to reduce the burden on eBird reviewers from a user base that is rapidly growing and increasingly made up of novices (J. Fitzpatrick, pers. comm. to UT admin). Employing more algorithms to use biased eBird data to inform eBird expectations, which subsequently bias eBird data, might only serve to fully automate a downward spiral of bias, rendering eBirders, eBird reviewers, and ecological reality, entirely irrelevant to the eBird database.

4.4 | Consequences of irreparably biased data

What consequences are there for researchers who assume eBird is a source of scientific data? At the smallest of scales, anyone viewing eBird data cannot know if the purported presence or absence of any species, at any location, at any time, represents reality or simply reflects when and where eBird accepted or rejected an unverifiable anecdotal report of that species. When the eBird database indicates a species was present at a location, was the species actually present, or was it in fact absent, but erroneously reported, and immediately accepted with no scrutiny?

When the eBird database indicates a species was absent from a location, was the species actually absent, or was it in fact present, correctly identified and reported, but rejected for refuting expectations? In either case, it is not possible for the researcher, or even eBird, to know the truth.

At larger scales, any analysis of eBird data regarding species' spatio-temporal ranges (Figure 2), or changes to those ranges (Figure 3), is almost certain to find patterns; however, those patterns are more likely artifacts of eBird's data collection, quality control, and curation processes than they are representations of reality. Additionally, although we did not include counts (i.e., reporting the number of individuals of each species observed) in our study, there are also eBird filters that set maximum numbers for a species that eBird expects to be present at any time and location. Therefore, with respect to abundances of birds, or changes in abundance over time, the same cycle of bias is at play. This might contribute to fantastically unrealistic species abundance estimates using eBird data, even in highly birded areas, contrary to Robinson et al. (2022). For example, Callaghan et al. (2021) used eBird data and estimated a global population of 541 Ivory-billed Woodpeckers (*Campephilus principalis*), a species for which uniquely great search effort has occurred in recent decades, with a realistic count of zero (Troy and Jones 2022).

Even if eBird data produced an unbiased estimate of a species' range, when that range actually contracts or shifts, eBird will continue to expect, and therefore accept without scrutiny, all erroneous reports of that species in the newly unoccupied area (Figure 3, Figure S1). Therefore, the only way a real range contraction can be accurately reproduced using eBird data is if eBird's *a piori* expectation of the species range was accurate and eBird adjusted those expectations concurrently with the contraction of the range. The same problem applies to temporal reductions or shifts in the period during which a species is present in an area, but those problems are likely magnified by the eBird temporal filters having a resolution of one week (J. Fitzpatrick, pers. comm. to UT admin). Additionally, if a true species' range expands or shifts into a new area, and eBirders submit true observations from the newly occupied area, eBird will reject 71–100% of those true observations, depending on how far the range has moved. Therefore, there may be only two ways that a real range expansion can be detected using eBird data. First, if the species is easy to identify (i.e., rarely erroneously reported) and the range expansion occurs within eBird's expected range (i.e., eBird expectations do not truncate the data), then real range expansions may be observable in eBird data (e.g., likely the case in Clark 2017). The second way, as with range contractions, is if eBird's *a priori* expectation of the species range is accurate, and eBird adjusts those expectations concurrently with the true expansion of the range. Also as with range contractions, the same problem applies to a lengthening or shift in the period during which a species is present in an area, which again is impacted by the one-week resolution of eBird's temporal filters. Furthermore, even if a history of eBird's spatial and temporal filter adjustments was available, the results of those adjustments cannot be reconstructed due to the annual alteration of 10% of eBird's historical content and the absence of data backups (see SI).

Similarly, any analysis of migration phenology, or changes thereto over time, even using percentage or proportion of birds passing through an area, is compromised by 1) the truncation of an immeasurable quantity of true early or late observations, 2) the blind acceptance of an immeasurable quantity of false but expected observations, 3) changes in expectation filters over time altering treatment of new and historic data, 4) the one-week temporal resolution of eBird filters, and 5) alternations to 10% of historical data every year with no backups. Even if changes detected using eBird data seem to corroborate reasonable hypotheses, like ranges shifting poleward or migration phenology advancing with climate change, those results are more likely a reflection of filter adjustments made within eBird based on expectations of species responses to climate change, in yet another form of confirmation bias. The confirmation bias in eBird reaches its pinnacle in analyses conducted by CLO eBird associates who "improve" analyses by excluding data from eBirders deemed to have low expertise because they submit too few expected species (Kelling et al. 2015, Johnston et al. 2018). Given the forced assimilation nature of eBird data, it is likely that eBird's eye-catching Spatio-Temporal Exploratory Models (STEM, ebird.org), produced with a $10M NSF grant, would be largely indistinguishable if eBird was fed observations ranging from entirely accurate to entirely random.

4.5 | Inviting trouble

A simple but important finding from our study is that eBird apparently does not verify users (i.e., nobody noticed our 196 new accounts reporting from a few IP addresses). It is free and simple to create an eBird account and submit information with no experience, no training, and no verification that the user ever looks at birds or is even human. Adding Terms of Use to a website, as eBird did for the first time in 2019, or requiring email verification upon account creation, does not protect a website from malicious users or bots (i.e., autonomous programs on the Internet that can interact with systems or users). This may seem trivial on its face to many ecologists. But in 2025, bots accounted for 53% of all online activity (Thales 2026). Just as bots create and use email and social media accounts (Ferrara et al. 2016), they also participate in internet-based scientific research (Pozzar et al. 2020, Griffin et al. 2022, Loebenberg et al. 2023). Computer scientists and social scientists have long recognized the threat to research data quality from bots and individuals misrepresenting themselves online (e.g. Prince et al. 2012), and they generally recommend investigators use comprehensive screening processes to reduce the impact (e.g., Bybee et al. 2022). eBird appears to do the opposite (see SI). It would be naïve to assume that bots and untruthful users do not make up a meaningful proportion of the >930,000 eBird accounts (ebird.org) and users of other web-based citizen science projects, while they make up large proportions of other web-based research projects that actually screen for them (Bybee et al. 2022). The recent addition of AI screening service Anibus (Techaro) demonstrates eBird's acknowledgement that bots regularly infiltrate their system. However, bots already bypass such protections (NIST 2025) as well as reCAPTCHA (i.e., prove you aren't a robot) screenings (Plezner et al. 2024), so such efforts do little to improve the integrity of the system.

Unfortunately, in addition to not screening users, many of the ways eBird and other citizen science projects incentivize participation also incentivize humans, and likely bots, to participate untruthfully. Dishonest behavior is prevalent when humans are in competition with one another, and more so in recreational activities that lack consequences for cheating (Sailors et al. 2017). Citizen science participation is generally a recreational activity that can be highly competitive, especially when involving birdwatching. eBird feeds that competition by posting real-time rankings among eBirders in various categories while offering valuable prizes for submitting observations following specific guidelines, typically emphasizing quantity, not quality, of contributions. Recently, the competitive nature of eBird has entered the avian ecology job market, as some employers request links to applicant eBird profiles (the authors, pers. obs.), and the CLO recommends applicants tout their eBird experience in cover letters (https://ebird.org/news/coverletter2016). The presumption that eBird participation reflects research acumen is almost as unsafe as the assumption that no applicant would pad their eBird account with fake observations in a competitive job market.

Unlike our study that was ethically designed consistent with standard practices of the relevant research community, it is doubtful that opportunistic individuals and bots participating fraudulently in research for money, prizes, to influence results, or for personal entertainment, are concerned about ethics. Ironically, in its current state, the eBird database is probably not impacted by contributions from invalid accounts owing to the cycle of confirmation bias that would result in largely the same dataset regardless of what is contributed or by whom. However, if eBird was able to recover all original observations, re-evaluate them objectively, and remediate the bias that pervades its database, that process and all data collection moving forward would benefit from authenticating users. Notably, multiple COVID-19 papers were retracted from elite medical journals when it was discovered that the sources of data – tens of thousands of patient records compiled by a third party – could not be verified (Piller and Servick 2020). Scrupulous journal editors, and funding agencies impressed by project participation rates, might consider holding all subject areas to a similar standard.

4.6 | Conclusion

We hope that our introduction of a method (i.e., web-based auditing) to the ecological sciences as a tool for assessing other methods (i.e., web-based data collection and curation) inspires additional creative, ethical, external scrutiny of similar programs. There are many additional questions that can be addressed within web-based data collection programs. It is doubtful that other citizen science projects are actively verifying users or supporting independent scrutiny of their systems either. For most, this is likely born of the same scarce resources that necessitated and popularized citizen science in the first place, but that is certainly not the case for eBird. We did not expect to learn that eBird is so unscientific. In hind sight, we acknowledge that our objective of helping improve eBird was naïve and may have exposed our own susceptibility to prestige bias. We were further surprised to learn after our study that eBird associates have been repeatedly made aware of similar concerns, and many others, but have been dismissive (ex-eBird associate, pers. comm. to UT admin). The causes and symptoms of premature canonization of ideas propagandized by charismatic despots (Bauer 2004, Tourish and Willmott 2023); the short-term benefits and long-term damages to academic culture from organizational narcissism (Vargo 2023); and the rage with which certain personality types respond to constructive criticism (Rasmussen 2016), are complex areas of psychology and ethics relevant to our experience but outside the scope of this paper. It is our sincere hope that legitimate web-based citizen science projects continue to make valuable contributions to science and conservation and that they are adequately supported in efforts to improve and expand. In a world of accelerating environmental issues and increasingly limited funding for science, maximizing reliable knowledge requires the entire scientific community to actively distinguish between legitimate scientific endeavors of disinterested seekers of truth and the status-branded products of self-interested “knowledge monopolies and research cartels” (Bauer 2004), with an observable differentiator being openness to a central tenant of science that is communal critique.

References

Altwegg, R., & Nichols, J. D. (2019). Occupancy models for citizen-science data. Methods in Ecology and Evolution, 10, 8–21. https://doi.org/10.1111/2041-210X.13090

Bauer, H. (2004). Science in the 21st Century: Knowledge monopolies and research cartels. Journal of Scientific Exploration, 18, 643–660.

Bernardi, R. A. & Nash, J. (2023). The importance and efficacy of controlling for social desirability response bias. Ethics & Behavior, 33, 413–429. https://doi.org/10.1080/10508422.2022.2093201

Bonter, D. N., & Cooper, C. B. (2012). Data validation in citizen science: a case study from Project FeederWatch. Frontiers in Ecology and the Environment, 10, 305–307. https://doi.org/10.1890/110273

Braschler, B. (2009). Successfully implementing a citizen-scientist approach to insect monitoring in a resource-poor country. BioScience, 59, 103–104. https://doi.org/10.1525/bio.2009.59.2.2

Brown, E., & Williams, B. K. (2019). The potential for citizen science to produce reliable and useful information in ecology. Conservation Biology, 33, 561–569. https://doi.org/10.1111/cobi.13223

Bybee, S., Cloyes, K., Baucom, B., Supiano, K., Mooney, K., & Ellington, L. (2022). Bots and nots: safeguarding online survey research with underrepresented and diverse populations, Psychology and Sexuality, 13, 901–911. https://doi.org/10.1080/19419899.2021.1936617

Callaghan, C. T., Nakagawa, S., & Cornwell, W. K. (2021). Global abundance estimates for 9,700 bird species. Proceedings of the National Academy of Sciences USA. 118, e2023170118. https://doi.org/10.1073/pnas.202317011

Clark, C. J. (2017). eBird records show substantial growth of the Allen's Hummingbird (*Selasphorus sasin sedentarius*) population in urban Southern California. Condor 119, 122–130.

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011). Assessing citizen science data quality: an invasive species case study. Conservation Letters, 4, 433–442. https://doi.org/10.1111/j.1755-263X.2011.00196.x

Cummings, O. W., Nowakowski, J. R., Schwandt, T. A., Eichelberger, R. T., Kleist, K. C., Larson, C. L. & Knott, T. G. (1988). Business Perspectives on Internal/External Evaluation. In McLaughlin, J. A., Weber, L. J., Covert, R. W. & Ingle, R. B. (Eds.), Evaluation Utilization, 39, 59–74.

Deguines, N., Julliard R., de Flores, M., & Fontaine, C. (2012). The whereabouts of flower visitors: contrasting land-use preferences revealed by a country-wide survey based on citizen science. PLoS One, 7, e45822. https://doi.org/10.1371/journal.pone.0045822

Delaney, D. G., Sperling, C. D., Adams, C. S., & Leung, B. (2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions, 10, 117–128.

Dickinson, J. L., Zuckerburg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: challenges and benefits. Annual Review of Ecology, Evolution, and Systematics, 41, 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59, 96–104. https://doi.org/10.1145/2818717

Foster-Smith, J., & Evans, S. M. (2003). The value of marine ecological data collected by volunteers. Biological Conservation, 113, 199–213. https://doi.org/10.1016/S0006-3207(02)00373-7

Gaddis, S. M. (2018). An introduction to audit studies in the social sciences. In: Gaddis, S. (eds) Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Methodos Series, vol 14.

Gardiner, M. M., Allee, L. L., Brown, P. M. Losey, J. E., Roy, H. E., & Smyth, R. R. (2012). Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment 10, 471–476. https://doi.org/10.1890/110185

Ge, Y., Knittel, C. R., MacKenzie, D., & Zoepf, S. (2020). Racial discrimination in transportation network companies, Journal of Public Economics, 190, art104205. https://doi.org/10.1016/j.jpubeco.2020.104205

Gorleri, F. C., & Areta, J. I. (2022). Misidentifications in citizen science bias the phenological estimates of two hard-to-identify *Elaenia* flycatchers. Ibis, 164, 13–26. https://doi.org/10.1111/ibi.12985

Grade, A. M., Chan, N. W., Gajbhiye, P., Perkins, D. J., Warren, P. S. (2022). Evaluating the use of semi-structured crowdsourced data to quantify inequitable access to urban biodiversity: A case study with eBird. PloS One 17(11): e0277223. https://doi.org/10.1371/journal.pone.0277223

Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., & Halkitis, P. N. (2022). Ensuring survey research data integrity in the era of internet bots. Quality & Quantity, 56, 2841–2852. https://doi.org/10.1007/s11135-021-01252-1

Johnston, A., Fink, D., Hochachka, W. M., & Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. Methods in Ecology and Evolution, 9, 88–97. https://doi.org/10.1111/2041-210X.12838

Hurlbert, A. H., & Liang, Z. (2012). Spatiotemporal variation in avian migration phenology: citizen science reveals effects of climate change. PLoS One, **7,** e31662. https://doi.org/10.1371/journal.pone.0031662

Imperva, Inc. (2023). Imperva bad bot report. https://www.imperva.com/resources/reports/2023-Imperva-Bad-Bot-Report.pdf (accessed 19 July 2023).

Thales (2026). Bad bot report: bad bots in the agentic age. https://cpl.thalesgroup.com/bad-bot-report (accessed 19 July 2026).

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T., & Donald, P.F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. Diversity and Distributions, 22, 1024–1035. https://doi.org/10.1111/ddi.12463

Kassing, J. W. (2008). Consider this: A comparison of factors contributing to employees' expressions of dissent. Communication Quarterly, 56, 342–355. https://doi.org/10.1080/01463370802240825

Kelling, S., Fink, D., La Sorte, F. A., Johnston, A., Bruns, N. E., & Hochachka, W. M. (2015). Taking a ‘big data’ approach to data quality in a citizen science project. Ambio 44, 601-611. https://doi.org/10.1007/s13280-015-0710-4

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. Frontiers in Ecology and the Environment, 14, 551–560. https://doi.org/10.1002/fee.1436

Le Pochat, V., Edelson, L., Van Goethem, T., Joosen, W., McCoy, D., & Lauinger, T. (2022). An audit of Facebook's political ad policy enforcement. In Proceedings of the 31st USENIX Security Symposium, USENIX Security 22, 607-624.

Light, R. J., Singer, J. D., & Willett, J. B. (1990). By design: planning research on higher education. Cambridge: Harvard University Press.

Loebenberg, G., Oldham, M., Brown, J., Dinu, L., Michie, S., Field, M., & Greaves, F. (2023). Bot or not? Detecting and managing participation deception when conducting digital research remotely. Journal of Medical Internet Research, 10.2196/46523. https://doi.org/10.1186/ISRCTN64052601

NIST. (2025). CVE-2025-24369, National Vulnerability Database. National Institute of Standard and Technology. https://nvd.nist.gov/vuln/detail/CVE-2025-24369. Viewed January 2026.

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 25, 447–453. https://doi.10.1126/science.aax2342

Ortega, R. P. (2021). Black scientists gather to form communities and boost diversity in science. Nature Medicine, 27, 756–758. https://doi.org/10.1038/s41591-021-01315-8

Panchariya, N. S., DeStefano, A. J., Nimbagal, V., Ragupathy, R., & Yavuz, S. (2015). Current developments in big data and sustainability sciences in mobile citizen science applications. 2015 IEEE First International Conference on Big Data Computing Service and Applications, Redwood City, CA, USA, 202–212.

Perkins, D. J. (2020). Blind spots in citizen science data: implications of volunteer biases in eBird data. M.S. Thesis, North Carolina State University, USA.

Piller, C. & Servick, K. (2020). Two elite medical journals retract coronavirus papers over data integrity questions. Science Magazine [online], 4 June 2020 (updated 10 June 2020).

https://www.sciencemag.org/news/2020/06/two-elite-medicaljournals-retract-coronavirus-papers-over-data-integrity-questions (viewed July 2020).

Plesner, A., Vontobel, T, & Wattenhofer, R. (2024). Breaking reCAPTCHAv2, 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 1047-1056, doi: 10.1109/COMPSAC61105.2024.00142.

Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsky, J.A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. Journal of Medical Internet Research, 22:e23021. https://doi.org/10.2196/23021

Prince, K. R., Litovsky, A. R., & Friedman-Wheeler, D. G. (2012). Internet-mediated research: beware of bots. The Behavior Therapist, 35, 85–88.

Rasmussen, K. (2016). Entitled vengeance: a meta-analysis relating narcissism to provoked aggression. Aggressive Behavior, 42, 362–379.

Rose, C. & Beck, V. (2016). Eyewitness accounts: false facts, false memories, and false identification. Journal of Crime and Justice, 39, 243–263, https://doi.org/10.1080/0735648X.2014.940999

Reese, G. C. & Skagen, S. K. (2017). Modeling nonbreeding distributions of shorebirds and waterfowl in response to climate change. Ecological Evolution, 7, 1497–1513. https://doi.org/10.1002/ece3.2755

Rutter, J. D., A. A. Dayer, H. W. Harshaw, N. W. Cole, J. N. Duberstein, D. C. Fulton, A. H. Raedeke, & R. M. Schuster (2021). Racial, ethnic, and social patterns in the recreation specialization of birdwatchers: an analysis of United States eBird registrants. Journal of Outdoor Recreation and Tourism 35:100400.

Sailors, P. R., Teetzel, S., & Weaving, C. (2017). Cheating, lying, and trying in recreational sports and leisure practices. Annals of Leisure Research 20, 563–577.

Sauer, J. R., Pardieck, K. L., Ziolkowski Jr, D. J., Pardieck, K. L., Fallon, J. E., & Link, W. A. (2017). The North American Breeding Bird Survey, Results and Analysis 1966-2015. Version 2.07.2017 U.S. Geological Survey Patuxent Wildlife Research Center, Laurel, MD, USA.

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. British Medical Journal, 351:h3480. https://doi.org/10.1136/bmj.h3480

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J. W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W. M., Iliff, M. J., Lagoze, C., La Sorte, F. A., Merrifield, M., et al. (2014). The eBird enterprise: an integrated approach to development and application of citizen science. Biological Conservation, 169, 31–40. https://doi.org/10.1016/j.biocon.2013.11.003

Soares, L., K. L. Cockle, E. R. Inzunza, J. T. Ibarra, C. I. Miño, S. Zuluaga, E. Bonaccorso, J. C. Ríos-Orjuela, et al. (2023). Neotropical ornithology: Reckoning with historical assumptions, removing systemic barriers, and reimagining the future. Ornithological Applications, 125, duac046. https://doi.org/10.1093/ornithapp/duac046

Supp, S. R., La Sorte, F. A., Cormier, TA, Lim, M. C. W., Powers, D. R., Wethington, S. M., Goetz, S., & Graham, C. H. (2015). Citizen science data provides new insight into annual and seasonal variation in migration patterns. Ecosphere, 6, 15. https://doi.org/10.1890/ES14-00290.1

Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. Current Biology, 24, 671–676. https://doi.org/10.1016/j.cub.2014.01.069

Troy, J. R., & Jones, C. D. (2023). The ongoing narrative of Ivory-billed Woodpecker rediscovery and support for declaring the species extinct. Ibis, 165, 340–351. https://doi.org/10.1111/ibi.13144

Tourish, D., & Willmott, H. (2023). Despotic leadership and ideological manipulation at Theranos: Towards a theory of hegemonic totalism in the workplace. Organization Studies, 44, 1801–1824. https://doi.org/10.1177/01708406231171801

Vargo, E. J. (2023). Organizational narcissism as an adaptive strategy in contemporary academia. Journal of Academic Ethics, 21, 293–302.

Walker, J. & Taylor, P. D. (2017). Using eBird data to model population change of migratory bird species. Avian Conservation and Ecology, 12, 4. https://doi.org/10.5751/ACE-00960-120104

Ward-Paige, C. A., Mora, C., Lotze, H. K., Pattengill-Semmens, C., McClenachan, L., Arias-Castro, E., & Myers, R. A. (2010). Large-scale absence of sharks on reefs in the greater Caribbean: a footprint of human pressures. PLoS One, 5, e11968. https://doi.org/10.1371/journal.pone.0011968

Figure 1.

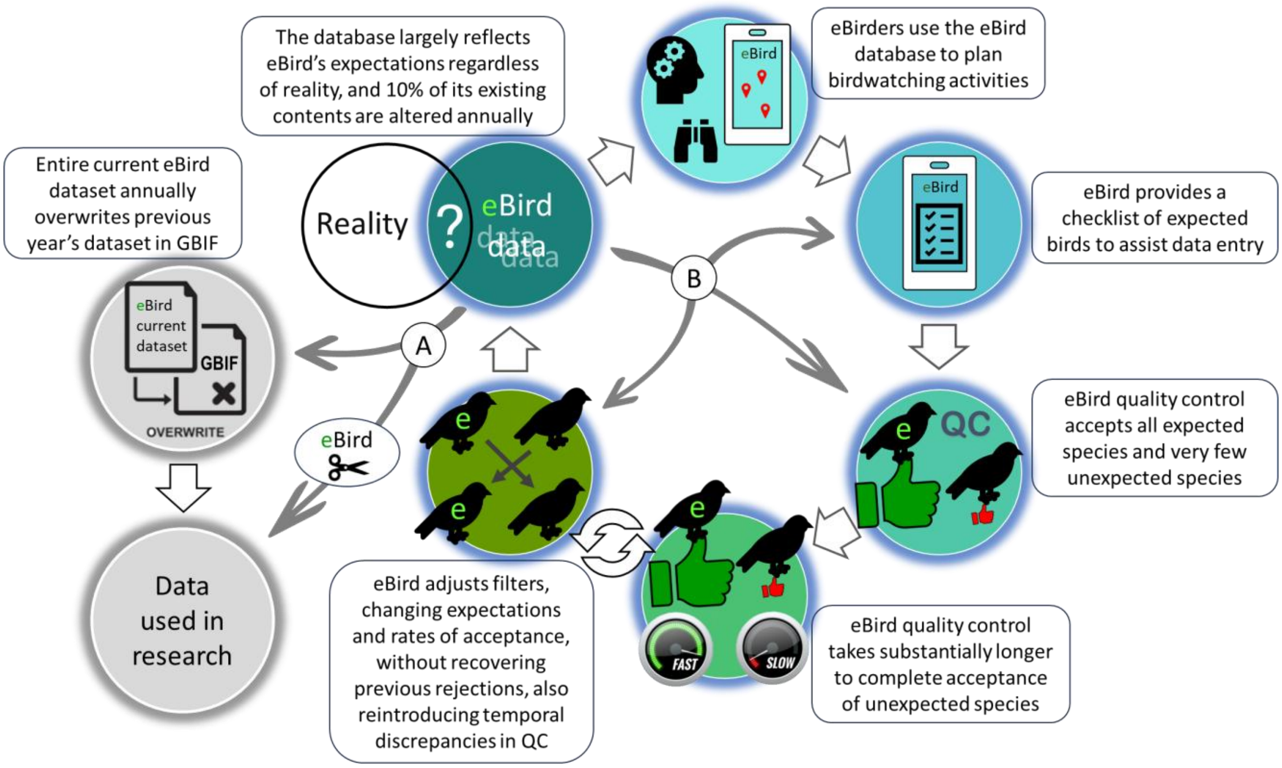

Figure 1. The compounding cycle of bias inherent to the design and implementation of eBird. Starting top center: Submissions from eBird users (eBirders) are non-independent when their birdwatching activities are informed by the eBird database; eBirder reports are compromised upon entry by a user interface that emphasizes species expected by eBird and reduces the probability of true but unexpected observations being reported; the eBird data quality control (QC) process selectively accepts (green thumb up) 100% of observations that eBird expects (bird with green "e") but accepts (red thumb up) very few observations they do not expect (solid black bird), regardless of observation validity; expected observations enter the eBird database immediately while the few unexpected observations that are accepted are delayed by weeks to years, increasing bias in recent data; eBird adjusts algorithm-based filters delineating when and where species are expected, with complex consequences for the database depend on whether and how quickly reviewers act following adjustment; the complete process results in a mercurial database of unverified anecdotes selectively filtered to reflect eBird's expectations regardless of reality, and for which 10% of historical data are altered annually with no backups maintained to track changes. The cycle of bias perpetuates itself when eBirders use the database to inform additional birdwatching activities. From this cycle, (A) data are extracted directly from eBird or from the Global Biodiversity Information Facility (GBIF) where eBird regularly deposits the full current database, overwriting the previous version, and leaving no backups. In the case of internal analysis by eBird associates, data are further biased (scissors) prior to analysis by prioritizing data based on eBirder "expertise" measured as proportion of expected observations submitted. Lastly, (B) the database that largely reflects eBird's expectations is used to inform adjustments to eBird's expectation filters, accelerating the spiral toward forced conformity of data and away from ecological reality.

Figure 2.

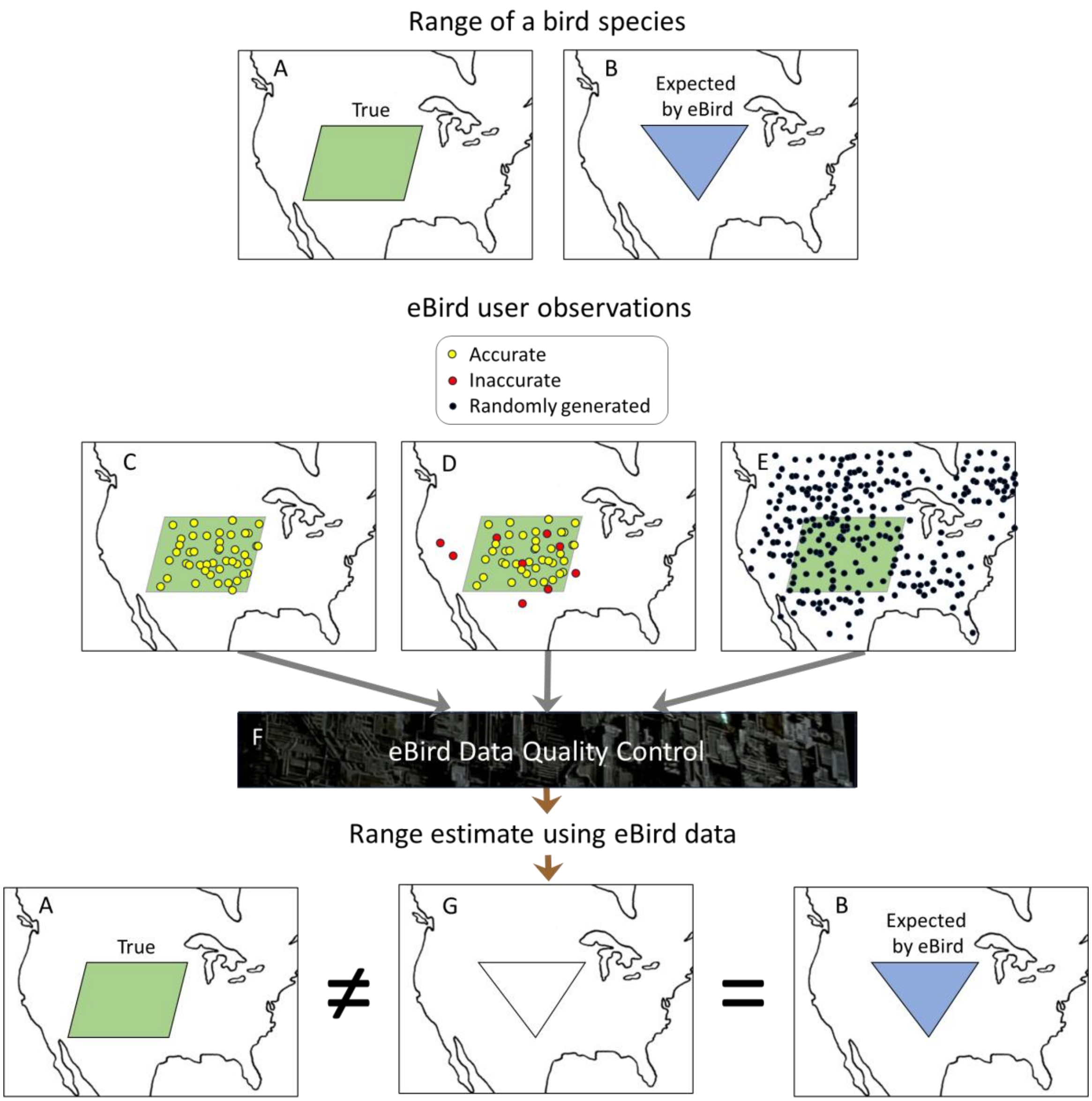

Figure 2. Confirmation bias imposed by the eBird data quality control process causes the true range of a species and eBirder observations both to be largely irrelevant to the eBird database and range estimates based thereon. Shown are (A) a hypothetical true range of a bird species in North America (green parallelogram); (B) hypothetical eBird *a priori* expectation of that species' range (blue triangle); dispersions of potential observations of the species reported to eBird, assuming (C) eBirder observations are 100% accurate, (D) eBirder observations are ~80% accurate, and (E) if eBirders were replaced by a random observation generator; (D) the eBird data quality control process that accepts 100% of observations within the expected range and rejects 71-100% of observations outside the expected range; and (G) the range that would be estimated by analysis of eBird data. One can replace space with time and find the same issue with temporal ranges of occupancy.

Figure 3.

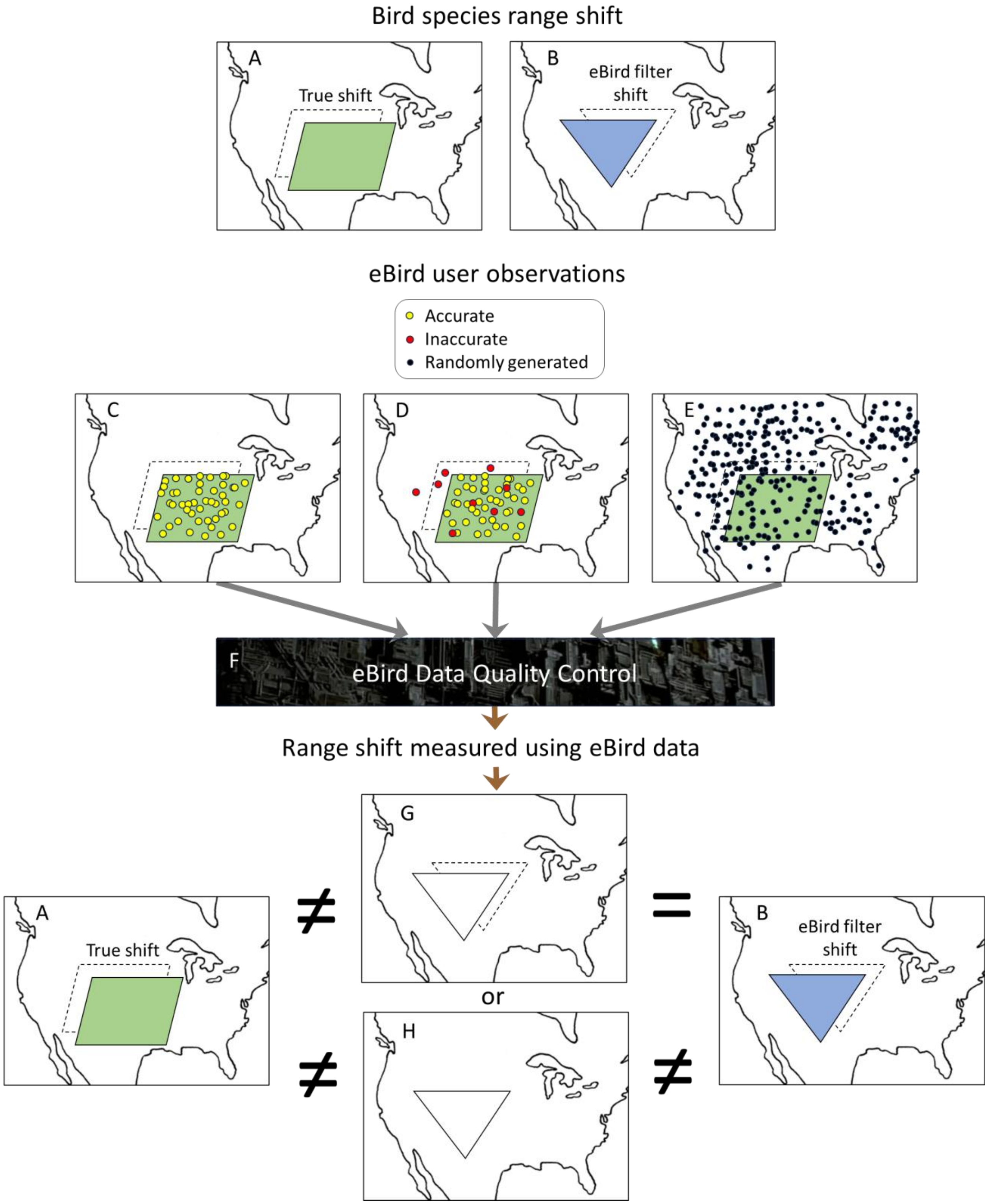

Figure 3. Confirmation bias imposed by the eBird data quality control process causes true changes in species ranges and eBirder submissions both to be largely irrelevant to the eBird

database and analyses based thereon. Shown are (A) a hypothetical true bird species range shift in North America (from dotted parallelogram to green parallelogram); (B) hypothetical change by eBird in their expectation of that species' range (from dotted triangle to blue triangle); dispersions of potential observations of the species reported to eBird, assuming (C) eBirder observations are 100% accurate, (D) eBirder observations are ~80% accurate, and (E) if eBirders were replaced by a random observation generator; (D) the eBird data quality control process that accepts 100% of observations within the expected range and rejects 71-100% of observations outside that expected range; and results that would be estimated by analysis of the eBird database depending on (G) if historically accepted observations that are outside of eBird's new expectations are re-accepted by eBird reviewers, or (H) they are not re-accepted and effectively lost from the database. The leading and trailing ends of these range shifts reflect range expansion and contraction, and thus these issues apply to any change in range. One can replace space with time and find the same issue with changes in temporal ranges of occupancy.

Supporting Information

Details on eBird account creation

In December 2016 and January 2017, we created 200 eBird accounts, later reduced to 196 accounts (see below). The creation of an eBird account requires an email address. Therefore, we created 200 Gmail (Google) accounts. We then used those email accounts to create the 200 eBird accounts for 200 simulated eBirders.

We assigned the simulated eBirders names that corresponded with first and last names socially perceived as being different sexes and races. From online lists of common names associated each group, we selected 25 first (or given) names in each of four groups socially perceived as White-sounding female, White-sounding male, Black-sounding female, and Black-sounding male. In addition, we selected 25 last names (or surnames) socially perceived as White-sounding and Black-sounding. We assembled the full names of the simulated eBirders by randomly combining the White-sounding first and last names and Black-sounding first and last names for males and females, yielding 200 names, 50 in each group. We confirmed that the names matched the desired perceived demographic groups by Google Image searching each name and observing the appearance of people in the first page of images. Our intention was to test for potential bias in the experience/treatment of eBirders of different perceived demographic groups, and not to represent the true demographics of eBirders because bias affecting minority citizen scientists would be difficult to detect with sample sizes reflecting real-world proportions (i.e., sample sizes of minorities would be too small).

We assigned four simulated eBirders, one from each demographic group, to each of the 50 United States. We later excluded Hawai'i because its unique avian system made it difficult to find locations and times to represent unexpected but plausible observations. To determine if there was geographic variation in acceptance rates of unexpected observations, we divided our simulated eBirders into groups based on geographic regions within the U.S., including Northeast (the New England states plus Delaware, Maryland, New York, New Jersey, and Pennsylvania), Midwest (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin), Pacific (Alaska, California, Oregon, and Washington), Rocky Mountain (Colorado, Idaho, Montana, Nevada, Utah, and Wyoming), Southeast (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia), and Southwest (Arizona, New Mexico, Oklahoma, and Texas).

It was possible that the inexperience of our simulated eBirders (i.e., new to eBird) could have subjected our reported bird lists to different levels of scrutiny from eBird than those of eBirders with long records of data submission, making our results potentially non-representative of eBird data acceptance and rejection rates in general. Although testing for experience-based bias would be an interesting area of inquiry, it was not our objective. Therefore, we took two precautions to limit possible experience-based bias. First, we loaded lists of birds that each simulated eBirder had “observed” prior to joining eBird (i.e., a life list). From a list of the 200 most common species in the U.S., we randomly generated life lists of 50-100 birds and applied those lists to each simulated eBirder account to give them a baseline level of experience. Second, through each account we submitted 4 - 6 lists of bird observations so that all simulated eBirders would have a record of reporting expected observations before submitting an unexpected observation.

Details on submission of fictitious observations

During January - June 2017, we submitted lists of bird observations to eBird through each simulated eBirder account. We did not observe any of the birds we reported and we did not visit any of the sites from which we reported lists. We submitted each list as if the simulated eBirder had visited a common birding location near the largest city in the state to which the simulated eBirder account was assigned. We produced our lists by perusing recent (i.e., within the previous few days) reports from purportedly real eBirders and submitting a list that represented ~75% of the species from a real eBirder’s list at our chosen location. In each case, we replicated other aspects of the real eBirder’s report, including “protocol”, “party size”, “duration”, and “distance” (see eBird) to make our reports as unremarkable as possible. We followed this protocol and submitted 4 - 6 lists of expected bird observations for each simulated eBirder during January - May 2017. These lists of expected bird observations served two purposes. First, they provided a large sample size with which to examine the low rate at which expected bird observations are likely rejected by eBird. Second, they provided a baseline of experience intended to reduce potential bias against our simulated eBirders when we later submitted unexpected bird observations.

During April and May 2017, we submitted an additional list for each simulated eBirder including a single unexpected, but plausible species observation along with expected species selected following the protocol above. The unexpected observations were included in 195 submissions; we excluded from analysis the list including an unexpected observation from one of the simulated eBirder accounts after we accidentally had that simulated eBirder submit a list of observations in the wrong state. Each unexpected observation was of a species that was slightly outside of eBird's expected range for that species in space or time. For example, we submitted an observation of a Fox Sparrow (*Passerella iliaca*) in Delaware County, Ohio on May 9, whereas eBird's expected temporal range for that species at that location ends in the first week of May. See WebData 1 for date, location, and species of each unexpected observation. Each unexpected observation was immediately flagged (i.e., identified as unexpected) by eBird's filtering algorithm, at which point we were prompted to enter an additional note to support the observation. Based on discussions with ex-eBird reviewers, we developed an *a priori* set of four brief notes, that included terms common in eBirder entries, from which we randomly selected for each such case. Those four responses included 1) "Saw Clearly", 2) "Seen Clearly", 3) "Clearly Observed", and 4) "Saw Well." After entering this note, we submitted the list of observations including the unexpected observation, which should have then been available for additional scrutiny.

Details on responding to eBird inquires

We consulted with three ex-eBird reviewers to confirm our understanding of how the eBird review process proceeds. Behind the scenes (i.e., outside of the citizen scientist's view) all unexpected observations are added to a list available to be scrutinized by a local/regional eBird reviewer(s). eBird reviewers are individuals with presumably relatively superior knowledge of the birds in a region (e.g., an area within a state or an entire small state in the United States); there may be one or multiple reviewers responsible for a region; and reviewers receive no compensation for their service. As of the time of our study, reviewers also received no guidance on how to scrutinize observations and no training on how to interact with eBirders. Despite claims that eBirders will have an opportunity to explain every unexpected report to reviewers via email, reviewers have the authority to accept or reject any observation without communicating with the eBirder. Reviewers can inquire about an observation through an email to the eBirder. Such inquiries are often a standard form email but can include additional text from the reviewer. However, they all request additional information about the unexpected observation. All the simulated eBirder email accounts were set to forward messages to a master email account, which we monitored. When one of our simulated eBirders received such an inquiry, we ignored the form and content of the message, and we ignored the email address, name, and everything else about the message, and we responded with a standardized message designed to represent typical eBirder emails, as per conversations with ex-eBird reviewers. Our standardized response included two points:

1. A phrase or statement confirming that the bird was observed either at close range or for a meaningful duration (typically 1 or 2 minutes).
2. A brief description of morphological characteristics and/or behavioral characteristic that are considered identifying for the species. We paraphrased morphological and behavioral

descriptions of birds from the species accounts on Cornell University's All About Birds website (www.allaboutbirds.org/).

For example, had we reported a Golden-crowned Kinglet (*Regulus satrapa*) outside of eBird's spatio-temporal expectations, we would respond to an eBird inquiry with, "I saw this bird at close range. It had a tiny body and a golden crown."

Details on observation acceptance or rejection

On 10 October 2017 (4-5 months after the last simulated eBirder submissions), we tallied the expected and unexpected bird observations that appeared in the eBird database (i.e., accepted) or did not appear in the eBird database (i.e., rejected). The appearance of an observation in the database confirmed that it was accepted either by the eBird filtering algorithm (expected observations) or by an eBird reviewer after being flagged as unexpected by the filtering algorithm. We then re-checked the status of each unexpected observation after another year (i.e., Oct. 2018), and again in January 2020 when we removed all observations from all simulated eBird accounts. No additional changes occurred to the status of any observations in that time. We could not determine if an observation excluded from the database was rejected, never addressed by a reviewer, or still awaited consideration >2.5 years after submission, but the resulting exclusion of the observation from the database had the same effect (i.e., rejection) regardless of the reason.

Safeguards implemented to minimize impact of intervention

Following audit research standard ethics (e.g., Sandvig et al. 2014), we designed our study to reduce, and in this case eliminate, possible influence of our intervention on anyone else interacting with the audited entity (i.e. eBird). First, our fictitious observations were of many different species and spread across a large geographic area over several months, such that they could not discernably contribute to the existing error rate among the millions of concurrent eBirder submissions. Second, we assigned the count of birds observed to “X” instead of reporting a number, so that no analyses of bird abundance would include our fictitious observations as data. Third, >99% of our fictitious observations were similar to real eBirder observations from similar dates and locations, so they could not impact any study of spatio-temporal occupancy. Fourth, our unexpected observations were very few, represented many species across a wide area, and were adjacent to eBird’s expectations for time and location, such that they could not impact analyses of species’ ranges. Lastly, we removed the simulated observations after the completion of the study, and there is subsequently no evidence they ever existed because eBird does not keep backups to enable recovery of data deleted by users, whether intentionally or accidentally (ebird.org). Moreover, the CLO over-writes their regular submissions to biodiversity repositories with the current version of the eBird database, which permanently eliminates any data removed between deposits (S. Kelling, CLO, pers. comm. to UT admin).

eBird is not a stable data repository

A central tenant of science is reproducibility. This requires detailed reporting of data-collection and analysis methods so studies can be replicated from scratch. It also helps when raw data are made available so analyses can be replicated, scrutinized, and expanded upon with new statistical methods. A troubling feature of the eBird database is that 10% of its contents are altered annually (S. Kelling, CLO, pers. comm. to UT admin). These changes are separate from, and in addition to, new current data being submitted each year. The annual alterations to historical data result from actions including but not limited to eBirders entering old data for the first time, eBirders editing or deleting existing checklists, eBird's delayed acceptance of old unexpected observations, and new changes in eBird's expectation filters that subsequently affect historical data. Therefore, eBird is not a stable data repository, and authors should not refer readers to eBird in data availability statements. Any attempt to extract the same dataset twice should be expected to produce differences that increase with the length of time between extractions. Additionally troubling, the CLO annually, or possibly quarterly (J. Fitzpatrick, pers. comm. to UT admin), deposits the entire current eBird database in the Global Biodiversity Information Facility (GBIF), which overwrites the previous year's submission, and thus eliminates any opportunity to review previous versions of the database.

How many eBirders are there?

The CLO reports that 930,000 eBirders from every country in the world have contributed more than 1.6 billion bird observations to eBird, and "average participation growth rate of approximately 20% year over year," (ebird.org, 5 January 2024). Presumably, that is the total number of eBird accounts ever created, including those created during beta testing, those created by bots, those that have never contributed bird observations, those that are otherwise inactive/dormant because the user started a new account, combined multiple accounts, or quit using eBird for any reason. This is because it was not possible, in the first two decades of eBird's existence, for an eBirder to delete an eBird account. A user could stop using their account, unsubscribe from emails, and even delete all the bird observations they had submitted, but their account could not be deleted, and thus remains in the total count of eBirders. It is not possible to know from publicly available records how many real/active eBird accounts exist despite this being presumably easy for the CLO to quantify and being important information for advertisers (greater participation is better), funding agencies (greater participation is better), and researchers (accurate counts of users is basic metadata).

> ## Can I delete my eBird account?
> It is not possible at this time to delete your eBird account. If you wish to no longer receive communication from the Cornell Lab, please click 'unsubscribe' at the bottom of the email that you receive (see Subscribe/Unsubscribe below).

Screenshot of eBird support webpage from 22 August 2022 (accessed via www.archive.org on 28 January 2024) indicating that it remained impossible to delete an eBird account at that time.

Only recently (early 2023), did the CLO allow users to delete eBird accounts. However, the CLO strongly disincentivizes eBird account deletion by tying each user's eBird account to all other Cornell services, including paid subscriptions.

> ## Delete Your Account
> If you wish to no longer receive email communications, see Subscribe/Unsubscribe above.
>
> By deleting your Cornell Lab account, **you will lose your data, life lists, media, courses, and subscriptions for ALL Cornell Lab projects**—this includes your eBird AND Merlin data, and any courses or subscriptions from Bird Academy and Birds of the World purchased by that account. **Deleted accounts cannot be recovered.**
>
> Deleting an account is rare, and deleted accounts cannot be recovered.
>
> If you wish to permanently delete ALL of your Cornell Lab account information, **click here to request account deletion.** This page can also be accessed by tapping your username in the upper right corner of the eBird website and selecting 'Cornell Lab Account'.

Screenshot of eBird support webpage from 1 May 2023 (accessed via www.archive.org on 28 January 2024) after the CLO began allowing eBird accounts to be deleted. Note that an eBird account still cannot be deleted without deleting the user's entire Cornell Lab account and losing access to paid subscriptions and other services.

Paradoxically, one can cancel their subscription to other CLO products, such as Birds of the World, while having no impact on their eBird account. Why would the CLO make it uniquely difficult and costly for users to delete their eBird accounts? As indicated in the main text, researchers, journal editors, advertisers, and funding agencies should insist on the reporting of real counts of real users when making decisions and forming opinions based on participation rates in any web-based research.

Statement regarding no results for regional or demographic bias

We carefully designed our study such that "The experimental plan does **not** fall within the purview of requiring review by an Institutional Review Board as per the Common Rule" (E. Bankert, Director of IRB, Dartmouth College, pers. comm. to UT admin; emphasis her own). Although eBird reviewers are humans, interacting with humans during research does not make them human research subjects if you learn nothing "about" those humans. This fact was acknowledged by a UT IRB Analyst prior to initiation of our study and again after completion of the study. This fact was also confirmed by a UT Investigation Committee who concurred with unanimous testimony from many experts including one in audit study ethics (C. Sandvig, University of Michigan). Unfortunately, shortly after UT received allegations of misconduct and threats of legal action from Cornell, the UT IRB demanded HMS destroy data gathered after any email was sent by us within the eBird review process, and to never speak of those data or results, and UT Admin has subsequently backed the IRB decision with the threat of discipline up to and including termination. IRBs are federally mandated entities, and their decisions cannot be overturned by anyone but themselves, and the UT IRB is obdurate. It was never our intention to leave the reader to speculate about those data and results based solely on the aggression with which the CLO fought to silence them.

Figure S1.

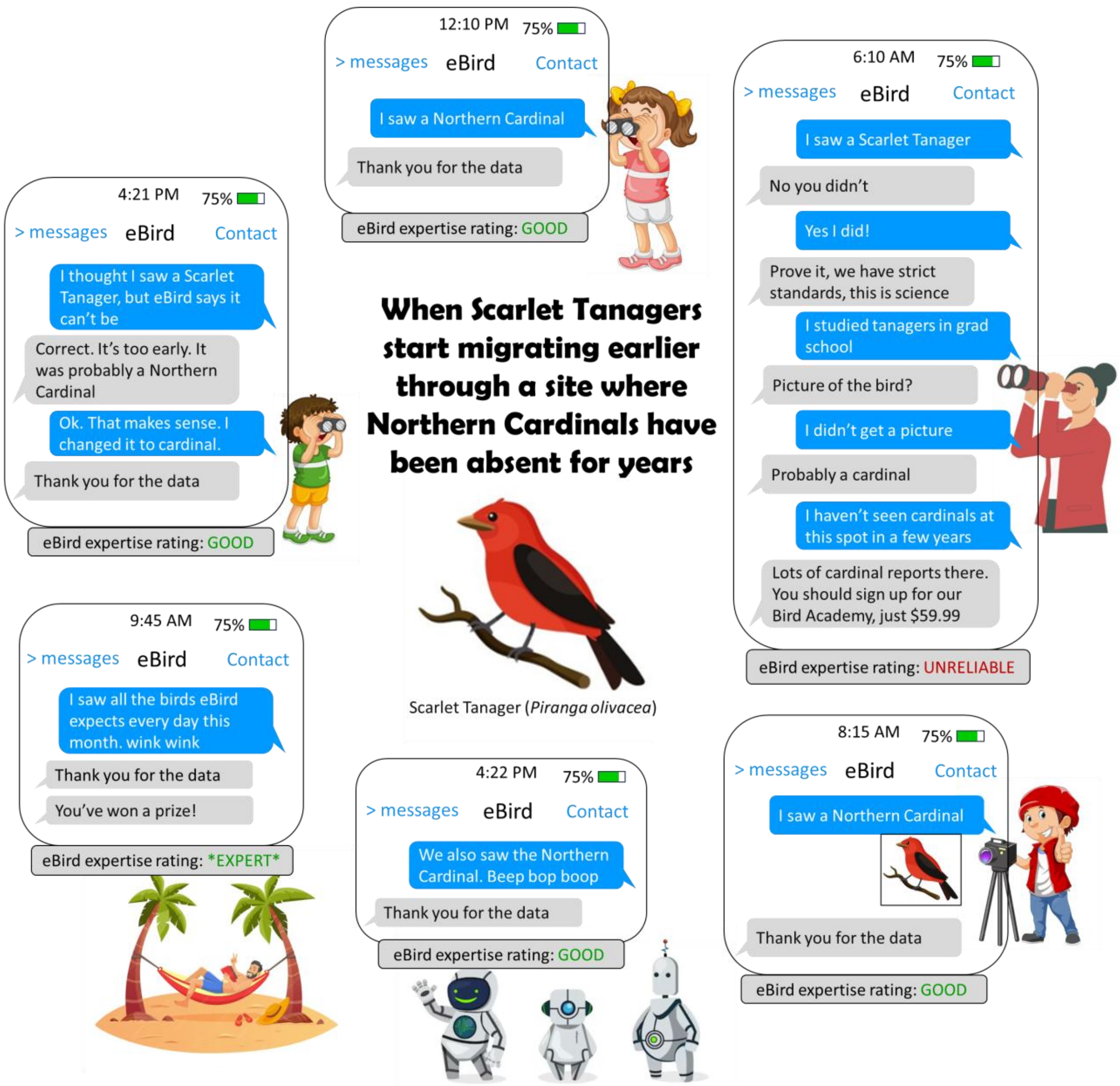

Figure S1. A schematic illustration of some of the issues associated with eBird's data collection methods. In this hypothetical scenario, migrating Scarlet Tanagers (*Piranga olivacea*) arrive in an area earlier than eBird's temporal filter is set to expect their arrival, while Northern Cardinals (*Cardinalis cardinalis*) have been absent from the area for multiple years but remain expected by eBird because eBirders continue to erroneously report them. Starting top center and moving counter clockwise: Any eBirder who erroneously reports a cardinal has their observation accepted with no scrutiny; Any eBirder who is uncertain as to which red bird they saw can be convinced by eBird's expected bird list and/or eBird reviewers that it was the expected cardinal; Any random person with an internet connection can submit anything from anywhere and have it accepted without scrutiny if it matches eBird's expectations, while eBird rankings and competitions incentivize such behavior, but eBird does not verify users; Some eBird accounts are likely bots, which make up more than half of internet activity and regularly participate in web-based research, but eBird does not verify users; Any eBirder that submits a photograph disproving their own observation will still have their cardinal report accepted with no scrutiny because it matches expectations and receives no scrutiny; Lastly, when an eBirder accurately identifies the unexpected tanager, there is a >70% chance that eBird will reject the observation; eBird then uses an "expertise" metric to rank the reliability of eBirders based on how many expected birds they report, and excludes low-expertise eBirder data from analysis. Note: this is an informational diagram; and eBird does not actually work via text message.