



OPEN

## Decision-making of citizen scientists when recording species observations

Diana E. Bowler<sup>1,2,3</sup>✉, Netra Bhandari<sup>1</sup>, Lydia Repke<sup>4</sup>, Christoph Beuthner<sup>4</sup>, Corey T. Callaghan<sup>1,5</sup>, David Eichenberg<sup>1,3</sup>, Klaus Henle<sup>6</sup>, Reinhard Klenke<sup>1,5</sup>, Anett Richter<sup>7</sup>, Florian Jansen<sup>8</sup>, Helge Bruelheide<sup>1,5</sup> & Aletta Bonn<sup>1,2,3</sup>

Citizen scientists play an increasingly important role in biodiversity monitoring. Most of the data, however, are unstructured—collected by diverse methods that are not documented with the data. Insufficient understanding of the data collection processes presents a major barrier to the use of citizen science data in biodiversity research. We developed a questionnaire to ask citizen scientists about their decision-making before, during and after collecting and reporting species observations, using Germany as a case study. We quantified the greatest sources of variability among respondents and assessed whether motivations and experience related to any aspect of data collection. Our questionnaire was answered by almost 900 people, with varying taxonomic foci and expertise. Respondents were most often motivated by improving species knowledge and supporting conservation, but there were no linkages between motivations and data collection methods. By contrast, variables related to experience and knowledge, such as membership of a natural history society, were linked with a greater propensity to conduct planned searches, during which typically all species were reported. Our findings have implications for how citizen science data are analysed in statistical models; highlight the importance of natural history societies and provide pointers to where citizen science projects might be further developed.

Citizen scientists—or volunteers contributing to scientific projects—increasingly take part in biodiversity monitoring by reporting species observations<sup>1,2</sup>. In Europe, for example, 87% of the participants in species monitoring are volunteers<sup>3</sup>. Chandler et al.<sup>2</sup> estimated over half of the data in the Global Biodiversity Information Facility (GBIF)—the largest global biodiversity database—comes from citizen science platforms. Species observations can be submitted by volunteers to a growing number of global platforms, such as iNaturalist and eBird, or regional platforms such as iRecord in the UK. The quantity of data is only set to increase as citizen science expands across people, places and taxa<sup>4</sup>. Citizen science data are also increasingly used in scientific research to address a broad range of questions about the large-scale patterns and long-term changes of biodiversity<sup>5–7</sup>. However, the quality of the data within citizen science platforms has been questioned<sup>8,9</sup>, potentially leading to a barrier of their widespread use<sup>10</sup>. There is a pressing need to better understand the heterogeneity within citizen science data to ensure that they are appropriately used in biodiversity research as well as conservation policy and practise.

Citizen scientists make many decisions—before, during, and after observing species—that can affect different aspects of the data that they collect and report. In large part, individual variability in decisions is constrained by the level of structure of the project. Citizen science projects range from unstructured (i.e., little to no training needed and few to no protocols to follow; e.g., iNaturalist) to semistructured (i.e., few protocols but important metadata collected as part of the recording process; e.g., eBird) to structured (i.e., a priori temporal and spatial

<sup>1</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, 04103 Leipzig, Germany. <sup>2</sup>Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany. <sup>3</sup>Department of Ecosystem Services, Helmholtz-Center for Environmental Research - UFZ, Permoserstraße 15, 04318 Leipzig, Germany. <sup>4</sup>Department of Survey Design and Methodology, GESIS - Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072 Mannheim, Germany. <sup>5</sup>Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle, Germany. <sup>6</sup>Department of Conservation Biology & Social-Ecological Systems, Helmholtz-Center for Environmental Research - UFZ, Permoserstraße 15, 04318 Leipzig, Germany. <sup>7</sup>Thünen Institute of Biodiversity, Bundesallee 65, 38116 Braunschweig, Germany. <sup>8</sup>Faculty of Agricultural and Environmental Sciences, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany. ✉email: diana.e.bowler@gmail.com

n = 899	Proportion (%) or median value (interquartile range)
Female	32%
Age (years)	55 (45–65)
Member of natural history society	42%
Number of years of experience	11 (4–30)

**Table 1.** Profile of the respondents.

sampling designs, often with trained volunteers; e.g., Reef Life Survey)<sup>11</sup>. In structured citizen science projects, participants follow a common protocol that guides many of the data collection decisions. However, the majority of species occurrence data come from unstructured citizen science projects, probably because of trade-offs in citizen science project design—the more protocols, the less participation<sup>11</sup>. In unstructured citizen science projects, individual observers make many independent decisions about, for instance, the focal types of habitat and species, the sampling times and methods, and survey durations. Variation among individual observers in these decisions, thus, creates heterogeneity in the data<sup>11–13</sup>. Moreover, in platforms compiling unstructured data, species occurrence records are usually not accompanied by any information about sampling methods and decisions<sup>14</sup>. Analysis of unstructured citizen science data often has to make many assumptions about how the data were collected. Violation of the assumptions of commonly used statistical models for unstructured data, such as occupancy detection models, can lead to inaccurate and/or imprecise predictions of the spatial or temporal patterns of species distributions<sup>15,16</sup>.

Many previous studies to understand the data collection decisions of citizen scientists have taken a data-driven approach by analysing the patterns in the available data<sup>4,14,17–22</sup>. Most of these studies focused on the spatial patterns of the data, for example finding evidence for higher sampling effort near human settlements and roads<sup>19</sup> as well as sometimes within protected areas<sup>20</sup>. Other studies have analysed observer-level patterns of species record submission to reveal the large variation among people in terms of recording intensity, taxonomic specialization, and preference towards rare species<sup>12,14</sup>. Data collected by citizen scientists have also been compared with those collected by professional surveyors to reveal differences in taxonomic identification skills and focus<sup>23–25</sup>. These studies indicate that common species are often underreported by citizen scientists, while rare species can be over-reported<sup>15,26–29</sup>.

An alternative approach to understanding citizen science data is by directly asking citizen scientists about their data collection activities. Interviews and questionnaires have been used to examine the motivations to participate in citizen science<sup>30–34</sup>, changes in conservation awareness and attitudes from participation<sup>35–39</sup>, and the types of people most likely to participate<sup>32,34,40,41</sup>. Questionnaires have also helped to understand some methodological aspects of data collection, including how citizen scientists interpret survey instructions<sup>23</sup> and study designs that promote continued engagement<sup>42,43</sup>. However, to our knowledge, there has not yet been a questionnaire study focused on understanding the data collection decisions of citizen scientists.

We developed a questionnaire to ask citizen scientists about all aspects of their decision-making when contributing a species observation to a citizen science platform or database. Our questions related to species observation methods, species and site selection, sampling effort, species identification uncertainty, as well as motivations to collect data and experience. We targeted our questionnaire towards people in Germany who voluntarily and independently collect species observations for unstructured citizen science schemes. We disseminated our questionnaire broadly to reach people who varied in expertise and taxonomic focus. By characterizing the decision-making process of citizen scientists collecting biodiversity data, we aimed to: (1) quantify the greatest sources of variability among observers, which may need to be incorporated into biodiversity models of unstructured data<sup>16</sup>; (2) examine whether variables related to experience and motivation could be used as proxies for data collection variation<sup>44</sup>, and (3) identify potential metadata that describes some of the variation among people and could be more routinely collected by citizen science data platforms<sup>45</sup>.

## Methods

**Participant recruitment.** The questionnaire was open from 15<sup>th</sup> October to 30<sup>th</sup> November 2020 and disseminated via multiple streams: personal contacts, natural history societies, German citizen science biodiversity platforms (Naturgucker/Naturwerke), Facebook groups and Twitter feeds of natural history groups. The questionnaire cover page explained the target audience and rationale for the questionnaire. The target audience was explained as people who voluntarily collect observations of plants or animals in their spare time, outside of a large-scale standardized monitoring program, and report these observations to an authority or organization. People of all levels of experience were encouraged to participate, from beginners to occasional collectors to experienced observers and experts. Respondents received no financial reward for participation.

**Participants.** Our analysis was based on a convenience sample of 1,645 individuals who originally viewed the questionnaire. Out of these individuals, 15% did not answer any questions, and 20% did not get beyond question 3. Our final sample included 899 respondents who reached at least the penultimate page of the questionnaire—out of which only 12 did not complete the final page requesting demographic information. This participation rate is comparable to established thresholds for response rates in survey research<sup>46</sup>. The majority of the respondents identified as male (67%) and were aged 45–65 years old (Table 1; Fig. S1). Geographically, respondents were widely spread across our study region of Germany (Fig. S2).

**Questionnaire description.** We designed the questions in accordance with methodological standards for surveys in the social sciences<sup>47,48</sup>. The title page of the questionnaire explained the rationale of the study and that participation in the questionnaire was voluntary. The questionnaire was approved as anonymous and not collecting any personal data by the legal department of the Helmholtz-Zentrum für Umweltforschung GmbH - UFZ and performed in accordance with relevant guidelines and regulations. The questionnaire was piloted on a pre-selected group of target participants and revised following their feedback.

Broadly, the main questions of interest could be grouped into the following nine sections.

- (1) Experience: participants were asked questions on the number of years collecting data and frequency of data collection, and, in a later section, on membership of natural history societies, formal knowledge of biodiversity monitoring and participation in any large-scale structured monitoring schemes.
- (2) Motivations: participants were asked to rate the importance of ten different aspects about why they record biodiversity on a 5-point Likert scale ranging from 'not important at all' to 'very important'. Our selection of motivation factors was guided by similar ones included in other studies, including both intrinsic factors (motivated directly by enjoyment of the activity) and extrinsic factors (motivated for reasons outside of enjoyment of the activity itself)<sup>49</sup>. For instance, we included 'have fun exploring' as an intrinsic motivation and 'support conservation' as an extrinsic motivation.
- (3) Survey types: participants were asked to report what proportion of their species observations come from different species survey types: active and planned species surveys (i.e., going to a place with the intention of looking for species), opportunistic observations not seen during an active search or observations made using traps—on a 5-point Likert scale ranging from 'none' to 'all'.
- (4) Active searches: participants were asked to rate the frequency with which they reported different kinds of species (e.g., all observed species or rare species only) on a 5-point Likert scale ranging from 'never' to 'very often' during an active and planned search and how long they typically spend looking for species (answering in minutes or hours).
- (5) Opportunistic observations: participants were asked to rate the frequency with which different scenarios (e.g., observations of rare species or simultaneous observations of many species) triggered opportunistic observations on a 5-point Likert scale ranging from 'never' to 'very often'.
- (6) Trap use: if participants previously indicated that they used traps, they were asked to rate the frequency with which they reported different kinds of species collected in their traps on a 5-point Likert scale ranging from 'never' to 'very often' and how long the traps were left active (answering in hours or days).
- (7) Species ID uncertainty: participants were asked the frequency with which they dealt with uncertainty about the taxonomic identification in different ways (e.g., not report or guess), on a 5-point Likert scale ranging from 'never' to 'very often'.
- (8) Locations: participants were asked to rate how often they looked for species in different habitats (e.g., forests, grasslands) on a 5-point Likert scale ranging from 'never' to 'very often'.
- (9) Consecutive surveys: participants were asked to rate how likely they were to report seeing a species again in the same place according to different time-periods since the previous detection of the same species, on a 5-point Likert scale ranging from 'not at all likely' to 'very likely'.

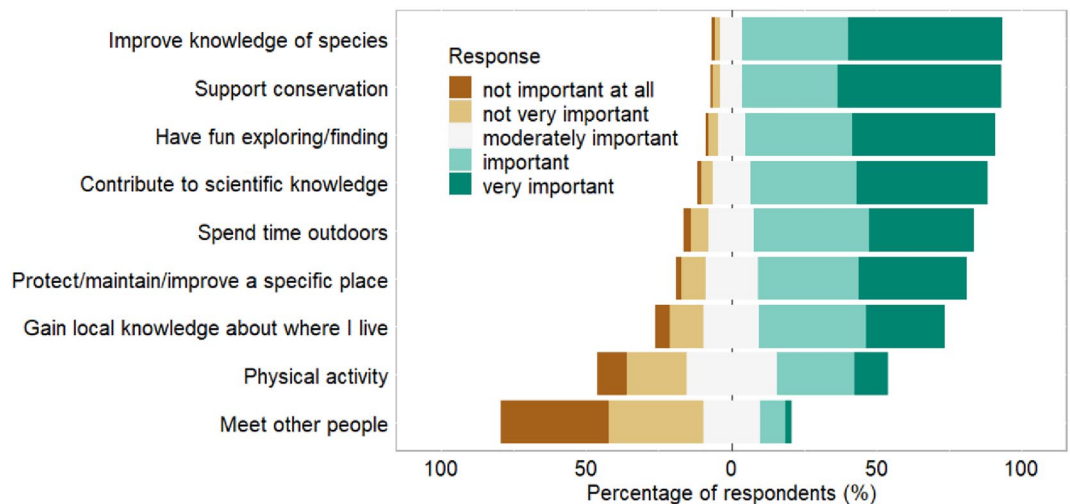
We asked participants to focus their answers with reference to their main taxonomic group of interest (options given: birds, plants, butterflies/moths, beetles, bees, dragonflies, amphibians/reptiles or other, to be specified). This was to simplify interpretation of the results, since people may survey differently for different taxon groups. We also asked people about their demographics (age, gender, location at a coarse postcode level), but this section was left optional in case of concerns about personal identification.

We organized and disseminated the questionnaire using the LimeSurvey web application. We used data-piping to ensure that questions remained logical according to responses to previous questions and to remind people of their selected focal taxon group within the question phrasing. We randomized the order of items within a question (e.g., in the question about motivation factors) to avoid any order effects, unless there was a logical order to the items. We also used validation to reduce mistakes in data entry (e.g., only allowing integers of maximum two characters for questions about number of years of experience). Finally, we ensured questions were specific with respect to a time-frame when necessary and kept the time frame recent (Spring and Summer 2020) to assume reasonable levels of recall. The questionnaire is available in "Appendix A" (original in German) and "Appendix B" (translation in English) and a summary of the items for each main question group is in Table S1.

**Statistical analysis.** Description: We conducted a descriptive analysis of the responses and visualized the numbers/proportions of people responding to each question with each answer option.

Linkages: We examined correlations among participant's responses across all main questions. We employed either polychoric, polyserial or Pearson correlations for the relationships between two ordinal variables, one ordinal variable and one continuous variable or two continuous variables, respectively. We excluded 'don't know' responses, which were consistently less than 5.5% of total responses for any item, and used the complete pair-wise data available for each correlation. We visualized the correlations using a chord diagram. Here, we were especially interested in the strength of the correlations between experience/motivations and data collection decisions.

Dimension reduction: We ran multiple analyses to assess whether responses to the questionnaire could be simplified to a smaller set of dimensions based on correlations among participants' responses to each question. We excluded questions about traps from these analyses since they were only applicable to a small proportion of the participants. We used the subset of data from respondents who did not respond 'don't know' to any question (n = 532 respondents). We first ran varimax-rotated principal components analysis (PCA) within each



**Figure 1.** Motivations of respondents to collect species observation data. Respondents were asked to rate the importance of each item. Items are ordered in the plot by the % responding ‘important’ or ‘very important’.

question group separately to identify the items that aligned most strongly with the first and second main axes of responses. Correlations among the top items of each question were then examined across question groups using varimax-rotated PCA and cluster analysis. For the cluster analysis, all values were scaled to between 0 and 1 and we used k-means partitioning, selecting the first three groups for simplicity of interpretation. All analyses and visualizations were done in R 4.1.0.

**Ethics declaration.** For informed consent by the participants, the title page of the questionnaire explained the rationale and purpose of the study and that participation in the questionnaire was voluntary. The questionnaire was approved as anonymous and not collecting any personal data by the legal department of the Helmholtz-Zentrum für Umweltforschung GmbH - UFZ and performed in accordance with relevant guidelines and regulations.

## Results

**Experience and motivations.** Respondents ranged between beginners and those with more than 50 years of experience collecting species observation data, but the median number of years of experience was 11 years (Table 1; Fig. S3). Most respondents collected species data at least on a weekly basis during spring or summer 2020 (Fig. S3). The most important motivating factors to collect data were supporting conservation (very important for 56.4% of people) and improving knowledge of species (very important for 53.2% of people) (Fig. 1). In contrast, some of the least important reasons for data collection included ‘physical activity’ (very important for 11.4% of people) and ‘meeting other people’ (very important for 2.1% of people).

Since our survey questions related to activity during spring/summer 2020 when people might have been restricted due to COVID-19 lockdowns, we also asked whether observation/reporting of species was different this year compared to previous years. Most people (64%) said that they had been unaffected in terms of data collection activity. Similar proportions of people said they were more (13%) or less (14%) active in 2021 compared with previous years. For the remaining 8%, this year was their first year.

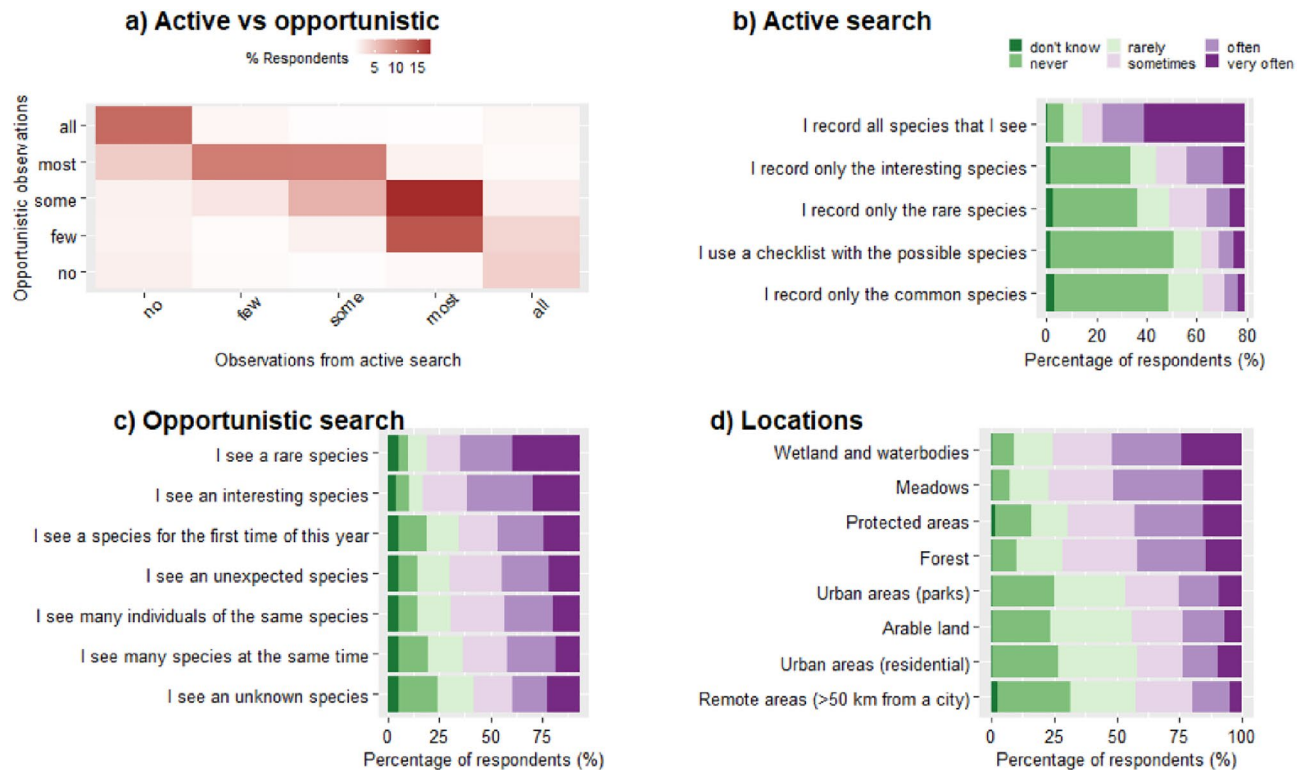
### Typical survey characteristics.

#### (a) Taxonomic foci

The most common focal taxonomic groups were birds (41%) followed by plants (18%), different insect groups (butterflies/moths, beetles, dragonflies/damselflies and bees) (19%, collectively), and amphibians/reptiles (9.2%) (Fig. S4). Most people collected data on any species within their identified focal group (Fig. S4). However, there were specialists, especially within the most species-rich groups, for instance, 40% of all beetle data collectors focused on specific beetle subgroups such as ground beetles (Carabidae) (Fig. S4).

#### (b) Types of surveys

In total, 45% of people reported that most or all of their species observations came from active/planned surveys; while 43% reported that most or all came from opportunistic observations. Typically, people reported to carry out a combination of active/planned surveys and opportunistic surveys (Fig. 2a). Traps were used by few



**Figure 2.** Survey patterns: (a) the proportion of species observation data that are made by an active/planned search compared with observations that were opportunistic; (b) species that are reported during an active/planned search; (c) triggers of an opportunistic observation; (d) locations/habitats in which people actively look for species.

recorders (<5% of respondents) focused on specific taxonomic groups; for instance, butterfly/moth specialists used different kinds of light traps (Table S2 for range of traps in use).

During an active/planned search, people stated that they reported most (38.5% of respondents) or all (14%) of the species that they saw (Fig. 2b). Consistent with this, few people reported only rare, only common or only species deemed interesting during active searches. Surveys were usually 2–3 h long, with some variation among taxonomic groups (Fig. S5).

#### (c) Triggers of opportunistic observations

When a species was reported after an incidental observation, observation of a rare species was the most common trigger (often or very often for over 50% of people). The least common triggers of opportunistic observations were related to species abundance and species richness, i.e., seeing many individuals of a species or seeing multiple species at the same time (Fig. 2c).

#### (d) Preferred locations

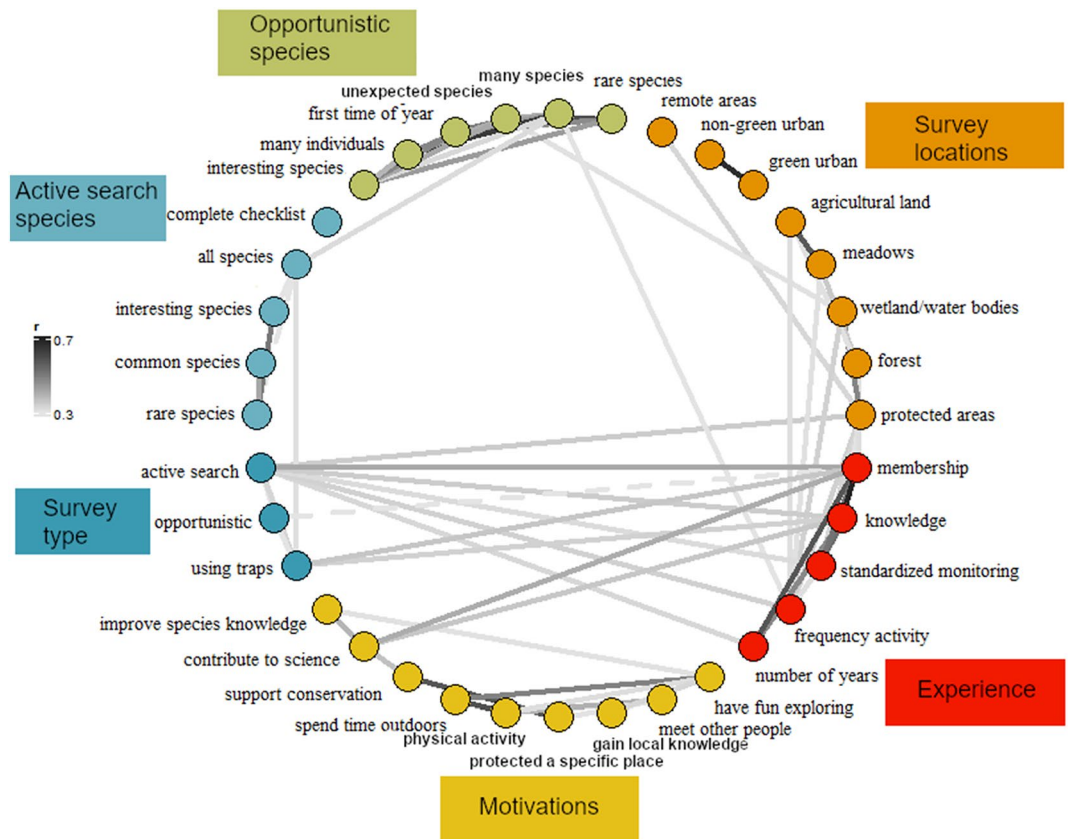
Respondents reported visiting different types of habitats to look for species, but with a tendency to more often visit semi-natural habitats than more human-modified habitats within urban or arable land (Fig. 2d). The most preferred places were open habitats such as meadows and wetlands/water bodies, which were visited often or very often by 52% of people. The least preferred habitats were arable land and urban areas (especially non-green), which were visited often or very often by 24% of people.

#### (e) Dealing with species identification uncertainty

When there was some uncertainty about species identification, most people used an identification guide to help (often or very often for over 84%). The least common approach was to simply guess the species identification (Fig. S6).

#### (f) Consecutive surveys

The reporting of a species observation on a given day depended on whether the person had previously reported the same species in the same location (Fig. S7). Most people were unlikely to record the same species



**Figure 3.** Plot showing the correlations among responses across all questions. Lines connecting two items indicate a correlation with absolute strength of 0.3 or greater between answers of each item, with darker shading indicating stronger correlation strengths. Negative correlations (very few) are shown by a dashed line. Table S1 gives a fuller description of all question items.

twice on the same day in the same place, but were increasingly likely to report it again with increasing time since the last observation: for instance, 25.2% were likely to report it again if seen in the previous week while 44% would report it again if seen in the previous year (Fig. S7).

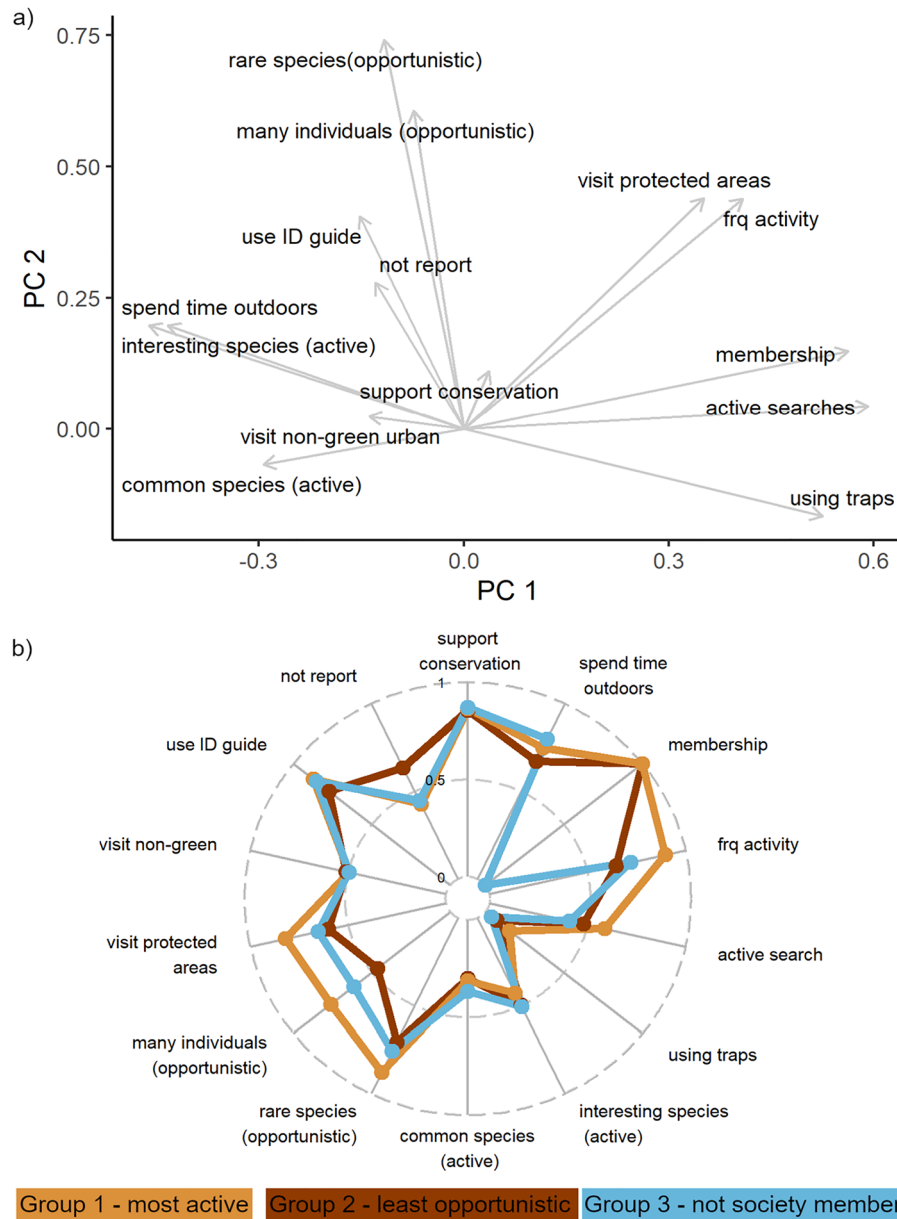
**Linkages across questions groups.** Motivations to report observations were not linked with any data collection methods or preferences, but there were some linkages between motivations and experience, albeit weak ones (Fig. 3). Contributing to science was an especially motivating factor for members of a natural history society ( $r=0.4$ ) and those who regard themselves as having knowledge of biodiversity monitoring ( $r=0.35$ ) (Fig. 3). Experience was also linked with some aspects of data collection, especially survey types and locations (Fig. 3). For instance, active/planned surveys were linked with experience attributes, especially membership of a natural history society ( $r=0.41$ ). Active/planned searches were also linked with visiting protected areas ( $r=0.34$ ). The proportion of records coming from active searchers was negatively correlated with the proportion coming from opportunistic observations ( $r=-0.7$ ), which is consistent with these search types being understood as mutually exclusive by the respondents. Frequency of activity was also linked with visiting multiple habitat types, including agricultural land.

**Dimension reduction to identify main axes of variation.** A PCA of responses to the top items (Table 2; Fig. S8) within all questions revealed that the first two components explained 25% of the variation among respondents. Natural history society membership and active/planned searches loaded more strongly on the first component (Fig. 4a). The second component was most associated with rare species as a trigger of opportunistic observations (Fig. 4a).

The best-defined three groups of respondents from a cluster analysis varied mostly in natural history society membership, propensity to conduct active searches, frequency of activity and triggers of opportunistic searches (Fig. 4b). Group 1 were the most active (in terms of frequency of activity) and were members of natural history societies and the most likely to conduct active searches. Group 2 were also consistently members of natural history societies but the least likely to report species opportunistically, and most likely not to report an observation when there was some uncertainty about species identification. Group 3 were not members of natural history societies and the least likely to conduct active searches. Groups were generally similar in terms of being motivated by supporting conservation.

Question group	Key items associated with variation among respondents
Experience	Society membership versus frequency of activity
Motivations	Spend time outdoors versus support conservation
Survey type	Active/planned search versus using traps
Active search species	Interesting species versus common species
Opportunistic species triggers	Rare species versus many individuals at the same time
Survey locations	Protected areas versus non-green urban areas
Species ID uncertainty	Use an identification guide versus not report

**Table 2.** Items explaining variation among people, as assessed by which items loaded most strongly onto each of the first two principal components of a PCA. Figure S8 shows the PCA biplots for each question group and Table S1 lists all items within each question group.



**Figure 4.** Dimension reduction: (a) a PCA analysis of the top items of all question groups (PC axis 1 explained 13% of the variation and PC axis 2 explained 12% of the variation) and (b) characteristics of the main respondent groups from a k-means cluster analysis. Points on each axis represent the mean value for people within each group (separated by different colours) scaled between the minimum and maximum values. Table S1 gives a fuller description of all question items.

## Discussion

Our study of citizen scientists across Germany demonstrates the diversity of decision-making and sampling methods underlying species occurrence records, consistent with previous studies analysing patterns in the available species data<sup>12,14</sup>. Nonetheless, by directly asking citizen scientists, our questionnaire approach allowed us to uncover some new patterns in the recording process of species observations. We found that: (1) volunteers were most often motivated by improving species knowledge and supporting conservation but motivations were not linked to data collection methods; (2) people often report observations from both opportunistic and active/planned surveys; however, the likelihood of complete surveys (i.e., reporting all species within their main focal taxon group) was highest during active/planned searches; (3) active/planned search were more typical of members of a natural history society; and (4) there was some temporal dependence in recording behaviour among consecutive surveys. Collectively, our findings have implications for how citizen science data are used in biodiversity research and highlight where citizen science projects might be further developed.

**Implications for using citizen science data in biodiversity research.** While various statistical methods have been proposed to account for the heterogeneity of unstructured citizen science data<sup>16,50,51</sup>, few studies have aimed to understand the data by directly asking citizen scientists about their data collection methods. In the absence of metadata on sampling methods, current approaches to model these data often involve creating proxies of survey type, effort and observer skill based on the available information<sup>50,51</sup>. For instance, many studies have calculated list length—the number of species reported on a given visit—and used this as a covariate to account for variation in survey type and effort<sup>51</sup>, assuming that recorders submitting more species were more likely to be performing a complete and/or longer survey. Similarly, other studies used proxies of observer experience, based on the total number of submitted records, to model variation in the data collected by different people<sup>44</sup>.

Our findings lend some support to these approaches but also reveal some further nuances. First, our questionnaire supported a contrast in the data coming from different survey types—specifically opportunistic observations compared with observations from active and planned searches. We found that the same people report observations from both opportunistic surveys and from active/planned surveys. We also found that people acted differently in each search type—reporting all species during active/planned searches but often only rare species opportunistically. Since rare species rather than groups of species at the same time were more typical triggers of opportunistic observations, a list length covariate may explain some, but not all, of the variation in survey types. Second, we found that experience-related variables were also linked with propensity to conduct active/planned searches, also supporting attempts to include experience proxies in analysis of unstructured data<sup>44</sup>.

Our data also revealed often overlooked patterns of data collection behaviour that have implications for analysis of citizen science data. For instance, we found that there was temporal dependence in recording behaviour. Specifically, people were less likely to report an observation of a species if they had already reported the species in the same place at previous time points. Methods to deal with such patterns have been suggested already, for instance, by borrowing methods developed to correct for trap-happiness found in some mark-recapture studies<sup>16</sup>, but they have rarely been applied in analyses of unstructured citizen science data so far. Studies of human preference decisions have tested the effects of novelty and familiarity<sup>52,53</sup>. These studies provide some support for a novelty preference within a nature setting, but at the same time highlight that this effect is context-dependent, for instance, depending on what task is being performed. However, such a novelty preference could explain the reduced likelihood of re-reporting a species at the same place as well as the increased likelihood to report a rare species opportunistically.

Since variation among people could not be reduced to just a few components, our findings are consistent with previous studies showing that people vary along multiple axes of recording behaviours<sup>12</sup>. Statistical methods might be able to account for some of this unmeasured heterogeneity using mixed-effects models with different random effects, e.g., for site and observer<sup>16</sup>, but some variation is likely to go unexplained since we show that observers themselves vary in their behaviour on different surveys. Hence, alongside developments in statistical methods, there should be better recognition of the importance of more detailed survey-level metadata to make the best use of unstructured citizen science data<sup>45</sup>.

**Increasing the value of data by more detailed metadata.** Lack of metadata associated with species occurrence records can lead to data heterogeneity being equated with low data quality. Hence, more detailed metadata could increase the utility of unstructured data for biodiversity research<sup>11</sup>. Some aspects of data collection can be characterized post hoc (e.g., survey habitat might be extractable using supplied geographic coordinates). Other metadata (e.g., time of day or survey duration) might be automatically identified by some software applications. However, other metadata needs to be provided manually by citizen science participants. Some platforms, such as eBird, already request that data submitters declare which kind of bird survey is associated with a set of observations, e.g., a complete checklist survey in which all observed species were reported<sup>45</sup>. Our analysis reveals the challenges of defining complete checklists for a broader range of taxa because complete checklists can only be defined with respect to a specific target taxonomic group. Some insect groups such as butterflies and dragonflies can represent target taxonomic groups since we found that few people specialized within these groups. By contrast, for other insect groups, especially species-rich groups such as beetles, the target taxonomic groups might be better defined at lower taxonomic levels, since here we found a larger proportion of specialists. Hence, metadata would need to be collected on both survey type (complete checklist or not) as well as target taxonomic group of interest.



**Implication for citizen science project design.** We found that people were motivated by both intrinsic and extrinsic factors<sup>49</sup>, but the top rated motivation factors were both extrinsic, suggesting that people are especially motivated by perceived benefits beyond enjoyment of the activity itself, especially the opportunity to increase species knowledge and support conservation<sup>54,55</sup>. To nurture long-term participation in citizen science, mechanisms could be put into place that provide feedback to participants on how their data are contributing to biodiversity research and conservation. Such mechanisms could lead to higher levels of long-term participation and retention in citizen science projects<sup>54,56</sup>. Citizen science participation has been linked with changes in attitudes towards conservation and conservation behaviours<sup>57</sup>). Since we found that motivations covaried somewhat with experience-related attributes, including membership of a natural history society, different projects and practices might encourage the involvement of participants with different backgrounds. Motivations did not consistently correlate with any data collection behaviour, hence, different types of citizen science projects can contribute useful data on species occurrence<sup>58</sup>.

Citizen science data is arguably most useful for ecological research by providing large spatial and temporal data sets that cannot be accomplished by academic scientists alone. Our analysis indicated that people most commonly look for species in open habitats, such as wetlands, water bodies and meadows<sup>13</sup>, and less often look in urban and agricultural areas. Despite being potentially less preferred, urban areas still are hotspots of data collection because of the large number of people living nearby<sup>19</sup>. However, agricultural areas, which are typically underrepresented in species occurrence record databases, and also less preferred as places to look for species according to our study, might need to be targeted by dedicated citizen science projects or standardized scientific surveys. Recent studies have developed methods to highlight under-sampled areas in order to encourage data collection and balance sampling effort across the landscape<sup>59</sup>.

**Key role of natural history societies.** We found that members of natural history societies had a greater propensity to carry out active/planned searches, which were associated with comprehensive surveys of all species, suggesting that their data might be especially useful for biodiversity research. Other studies have already shown the value of unstructured data from natural history groups for assessment of large-scale and long-term trends of insect taxa, which are mostly not targeted by large-scale structured monitoring schemes<sup>60,61</sup>. Natural history societies may thus serve as both an important mechanism for generating valuable biodiversity data, but also as a mechanism for recruitment of volunteers to contribute data and sharing of taxonomic knowledge and expertise. Natural history societies could be better supported in their needs for this role by both governmental agencies as well as research funders and invited to collaborate with professional scientists.

**Study limitations and caveats.** Although our study was able to reveal multiple aspects of citizen scientist decision-making, it had some particular aspects that might affect the generality of our findings. First, it is likely that due to the nature of our survey and the snowballing approach of dissemination, many of the respondents were probably the most active citizen scientists at the high end of recording frequency. Hence, our findings may not reveal the full breadth of recording patterns across citizen scientists. However, since most data are collected by a small number of highly active recorders<sup>12–14</sup>, our findings remain highly useful to understand recording activities underlying the data within big databases that are commonly used for ecological research. Second, we focused on the year 2020 to ascertain current behaviour, but recording activities might have been affected by the COVID-19 pandemic. As elicited by a specific question related to this, few people reported to have been affected; although, probably the most affected people did not participate in our questionnaire at all.

## Conclusions

Harnessing the full potential of citizen science data requires better understanding of how the data are collected. We demonstrated the heterogeneous nature of individual variability in citizen scientists, and their decisions as to when, how, and where to sample biodiversity. Improved metadata associated with species observations can help devise the most appropriate analysis for questions on biodiversity change. Studies of citizen scientists, such as ours, as well as studies of patterns in the available data<sup>12</sup> will help paint a better picture for how to extract the maximum information for biodiversity research.

## Data availability

Data are published in the iDiv data repository, <https://idata.idiv.de> (ID: 3512). Code to analyse the data is available at: [https://github.com/bowlerbear/Citizen\\_science\\_manuscript](https://github.com/bowlerbear/Citizen_science_manuscript).

Received: 8 December 2021; Accepted: 21 June 2022

Published online: 30 June 2022

## References

1. Fink, D. *et al.* Crowdsourcing meets ecology: the misperception of wide spatiotemporal species distribution models. *AI Mag.* **35**, 19–30. <https://doi.org/10.1609/aimag.v35i2.2533> (2014).
2. Chandler, M. *et al.* Contribution of citizen science towards international biodiversity monitoring. *Biol. Cons.* **213**, 280–294. <https://doi.org/10.1016/j.biocon.2016.09.004> (2017).
3. Schmeller, D. S. *et al.* Advantages of volunteer-based biodiversity monitoring in Europe. *Conserv. Biol.* **23**, 307–316. <https://doi.org/10.1111/j.1523-1739.2008.01125.x> (2009).
4. Boakes, E. H. *et al.* Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol.* <https://doi.org/10.1371/journal.pbio.1000385> (2010).
5. Follett, R. & Strezov, V. An analysis of citizen science based research: Usage and publication patterns. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0143687> (2015).

6. Zattara, E. E. & Aizen, M. A. Worldwide occurrence records suggest a global decline in bee species richness. *One Earth* **4**, 114–123. <https://doi.org/10.1016/j.oneear.2020.12.005> (2021).
7. Dickinson, J. L. *et al.* The current state of citizen science as a tool for ecological research and public engagement. *Front. Ecol. Environ.* **10**, 291–297. <https://doi.org/10.1890/110236> (2012).
8. Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **14**, 551–560. <https://doi.org/10.1002/fee.1436> (2016).
9. Bayraktarov, E. *et al.* Do big unstructured biodiversity data mean more knowledge?. *Front. Ecol. Evol.* <https://doi.org/10.3389/fevo.2018.00239> (2019).
10. Burgess, H. K. *et al.* The science of citizen science: Exploring barriers to use as a primary research tool. *Biol. Cons.* **208**, 113–120. <https://doi.org/10.1016/j.biocon.2016.05.014> (2017).
11. Isaac, N. J. B. & Pocock, M. J. O. Bias and information in biological records. *Biol. J. Lin. Soc.* **115**, 522–531. <https://doi.org/10.1111/bij.12532> (2015).
12. August, T., Fox, R., Roy, D. B. & Pocock, M. J. O. Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-67658-3> (2020).
13. Boakes, E. H. *et al.* Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Sci. Rep.* <https://doi.org/10.1038/srep33051> (2016).
14. Di Cecco, G. J. *et al.* Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *Bioscience* **71**, 1179–1188. <https://doi.org/10.1093/biosci/biab093> (2021).
15. Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T. & Donald, P. F. Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Divers. Distrib.* **22**, 1024–1035. <https://doi.org/10.1111/ddi.12463> (2016).
16. Altwegg, R. & Nichols, J. D. Occupancy models for citizen-science data. *Methods Ecol. Evol.* **10**, 8–21. <https://doi.org/10.1111/2041-210x.13090> (2019).
17. Courter, J. R., Johnson, R. J., Stuyck, C. M., Lang, B. A. & Kaiser, E. W. Weekend bias in citizen science data reporting: Implications for phenology studies. *Int. J. Biometeorol.* **57**, 715–720. <https://doi.org/10.1007/s00484-012-0598-7> (2013).
18. Amano, T., Lamming, J. D. L. & Sutherland, W. J. Spatial gaps in global biodiversity information and the role of citizen science. *Bioscience* **66**, 393–400. <https://doi.org/10.1093/biosci/biw022> (2016).
19. Geldmann, J. *et al.* What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers. Distrib.* **22**, 1139–1149. <https://doi.org/10.1111/ddi.12477> (2016).
20. Girardello, M. *et al.* Gaps in butterfly inventory data: A global analysis. *Biol. Cons.* **236**, 289–295. <https://doi.org/10.1016/j.biocon.2019.05.053> (2019).
21. Husby, M., Hoset, K. S. & Butler, S. Non-random sampling along rural-urban gradients may reduce reliability of multi-species farmland bird indicators and their trends. *Ibis* <https://doi.org/10.1111/ibi.12896> (2021).
22. Petersen, T. K., Speed, J. D. M., Grøtan, V. & Austrheim, G. Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. *Ecol. Solut. Evid.* <https://doi.org/10.1002/2688-8319.12048> (2021).
23. Egerer, M., Lin, B. B. & Kendal, D. Towards better species identification processes between scientists and community participants. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2019.133738> (2019).
24. Jimenez, M. F., Pejchar, L. & Reed, S. E. Tradeoffs of using place-based community science for urban biodiversity monitoring. *Conserv. Sci. Pract.* <https://doi.org/10.1111/csp2.338> (2021).
25. Branchini, S. *et al.* Using a citizen science program to monitor coral reef biodiversity through space and time. *Biodivers. Conserv.* **24**, 319–336. <https://doi.org/10.1007/s10531-014-0810-7> (2015).
26. Snall, T., Kindvall, O., Nilsson, J. & Part, T. Evaluating citizen-based presence data for bird monitoring. *Biol. Cons.* **144**, 804–810. <https://doi.org/10.1016/j.biocon.2010.11.010> (2011).
27. Gardiner, M. M. *et al.* Lessons from lady beetles: Accuracy of monitoring data from US and UK citizen-science programs. *Front. Ecol. Environ.* **10**, 471–476. <https://doi.org/10.1890/110185> (2012).
28. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* <https://doi.org/10.1038/s41598-017-09084-6> (2017).
29. Johansson, F. *et al.* Can information from citizen science data be used to predict biodiversity in stormwater ponds?. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-66306-0> (2020).
30. Everett, G. & Geoghegan, H. Initiating and continuing participation in citizen science for natural history. *BMC Ecol.* <https://doi.org/10.1186/s12898-016-0062-3> (2016).
31. Richter, A. *et al.* The social fabric of citizen science drivers for long-term engagement in the German butterfly monitoring scheme. *J. Insect Conserv.* **22**, 731–743. <https://doi.org/10.1007/s10841-018-0097-1> (2018).
32. MacPhail, V. J., Gibson, S. D. & Colla, S. R. Community science participants gain environmental awareness and contribute high quality data but improvements are needed: Insights from Bumble Bee Watch. *PeerJ* <https://doi.org/10.7717/peerj.9141> (2020).
33. Maund, P. R. *et al.* What motivates the masses: Understanding why people contribute to conservation citizen science projects. *Biol. Conserv.* <https://doi.org/10.1016/j.biocon.2020.108587> (2020).
34. Moczek, N., Nuss, M. & Kohler, J. K. Volunteering in the citizen science project “Insects of Saxony”—The larger the island of knowledge, the longer the bank of questions. *Insects* <https://doi.org/10.3390/insects12030262> (2021).
35. Branchini, S. *et al.* Participating in a citizen science monitoring program: Implications for environmental education. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0131812> (2015).
36. Kelemen-Finan, J., Scheuch, M. & Winter, S. Contributions from citizen science to science education: An examination of a biodiversity citizen science project with schools in Central Europe. *Int. J. Sci. Educ.* **40**, 2078–2098. <https://doi.org/10.1080/09500693.2018.1520405> (2018).
37. Deguines, N., Prince, K., Prevot, A. C. & Fontaine, B. Assessing the emergence of pro-biodiversity practices in citizen scientists of a backyard butterfly survey. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2020.136842> (2020).
38. Peter, M., Diekötter, T., Höffler, T. & Kremer, K. Biodiversity citizen science: Outcomes for the participating citizens. *People Nat.* **3**, 294–311. <https://doi.org/10.1002/pan3.10193> (2021).
39. Phillips, T. B., Bailey, R. L., Martin, V., Faulkner-Grant, H. & Bonter, D. N. The role of citizen science in management of invasive avian species: What people think, know, and do. *J. Environ. Manage.* <https://doi.org/10.1016/j.jenvman.2020.111709> (2021).
40. Parrish, J. K. *et al.* Hoping for optimality or designing for inclusion: Persistence, learning, and the social network of citizen science. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1894–1901. <https://doi.org/10.1073/pnas.1807186115> (2019).
41. Mac Domhnaill, C., Lyons, S. & Nolan, A. The citizens in citizen science: Demographic, socioeconomic, and health characteristics of biodiversity recorders in Ireland. *Citiz. Sci.: Theory Pract.* **5**, 16. <https://doi.org/10.5334/cstp.283> (2020).
42. van der Wal, R., Sharma, N., Mellish, C., Robinson, A. & Siddharthan, A. The role of automated feedback in training and retaining biological recorders for citizen science. *Conserv. Biol.* **30**, 550–561. <https://doi.org/10.1111/cobi.12705> (2016).
43. Bloom, E. H. & Crowder, D. W. Promoting data collection in pollinator citizen science projects. *Citiz. Sci.: Theory Pract.* **5**, 3. <https://doi.org/10.5334/cstp.217> (2020).
44. Johnston, A., Fink, D., Hochachka, W. M. & Kelling, S. Estimates of observer expertise improve species distributions from citizen science data. *Methods Ecol. Evol.* **9**, 88–97. <https://doi.org/10.1111/2041-210x.12838> (2018).
45. Kelling, S. *et al.* Using semistructured surveys to improve citizen science data for monitoring biodiversity. *Bioscience* **69**, 170–179. <https://doi.org/10.1093/biosci/biz010> (2019).

46. Koen, B., Loosveldt, G., Vandenplas, C. & Stoop, I. Response rates in the European social survey: Increasing, decreasing, or a matter of fieldwork efforts?. *Surv. Methods: Insights Field* <https://doi.org/10.13094/SMIF-2018-00003> (2018).
47. Gideon, L. *Handbook of Survey Methodology for the Social Sciences* (Springer, 2012).
48. Wolf, C., Joye, D., Smith, T. W. & Fu, Y. C. *The SAGE Handbook of Survey Methodology* (SAGE Publications Ltd, 2016).
49. Richter, A. *et al.* Motivation and support services in citizen science insect monitoring: A cross-country study. *Biol. Conserv.* **263**, 109325. <https://doi.org/10.1016/j.biocon.2021.109325> (2021).
50. Johnston, A., Moran, N., Musgrove, A., Fink, D. & Baillie, S. R. Estimating species distributions from spatially biased citizen science data. *Ecol. Model.* <https://doi.org/10.1016/j.ecolmodel.2019.108927> (2020).
51. Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P. & Roy, D. B. Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods Ecol. Evol.* **5**, 1052–1060. <https://doi.org/10.1111/2041-210x.12254> (2014).
52. Liao, H.-L., Yeh, S.-L. & Shimojo, S. Novelty vs. familiarity principles in preference decisions: Task context of past experience matters. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2011.00043> (2011).
53. Park, J., Shimojo, E. & Shimojo, S. Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14552–14555. <https://doi.org/10.1073/pnas.1004374107> (2010).
54. Tiago, P., Gouveia, M. J., Capinha, C., Santos-Reis, M. & Pereira, H. M. The influence of motivational factors on the frequency of participation in citizen science activities. *Nat. Conserv.-Bulg.* <https://doi.org/10.3897/natureconservation.18.13429> (2017).
55. Davis, A., Taylor, C. E. & Martin, J. M. Are pro-ecological values enough? Determining the drivers and extent of participation in citizen science programs. *Hum. Dimens. Wildl.* **24**, 501–514. <https://doi.org/10.1080/10871209.2019.1641857> (2019).
56. Bell, S. *et al.* What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. *Biodivers. Conserv.* **17**, 3443–3454. <https://doi.org/10.1007/s10531-008-9357-9> (2008).
57. Toomey, A. H. & Domroese, M. C. Can citizen science lead to positive conservation attitudes and behaviors?. *Hum. Ecol. Rev.* **20**, 50–62 (2013).
58. Dennis, E. B., Morgan, B. J. T., Brereton, T. M., Roy, D. B. & Fox, R. Using citizen science butterfly counts to predict species population trends. *Conserv. Biol.* **31**, 1350–1361. <https://doi.org/10.1111/cobi.12956> (2017).
59. Callaghan, C. T., Poore, A. G. B., Major, R. E., Rowley, J. J. L. & Cornwell, W. K. Optimizing future biodiversity sampling by citizen scientists. *Proc. R. Soc. B-Biol. Sci.* <https://doi.org/10.1098/rspb.2019.1487> (2019).
60. Outhwaite, C. L., Gregory, R. D., Chandler, R. E., Collen, B. & Isaac, N. J. B. Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nat. Ecol. Evol.* **4**, 384. <https://doi.org/10.1038/s41559-020-1111-z> (2020).
61. Bowler, D. E. *et al.* Winners and losers over 35 years of dragonfly and damselfly distributional change in Germany. *Divers. Distrib.* <https://doi.org/10.1111/ddi.13274> (2021).

## Acknowledgements

We thank all contributors of the online survey for their contributions and insights. We gratefully acknowledge the support of iDiv and the strategic project 'sMon' funded by the German Research Foundation (grant number DFG-FZT 118, 202548816). AB acknowledges funding by the German Ministry of Environment (BMU/BfN project VielfalterGarten). CTC was supported by a Marie Skłodowska-Curie Individual Fellowship (no. 891052). AR has worked for this publication within the framework of the nationwide project on monitoring of biological diversity in agricultural landscapes (MonViA) funded by the German Federal Ministry of Food and Agriculture (BMEL). Finally, we thank an anonymous reviewer for constructive comments.

## Author contributions

D.B. led the project, with input from A.B. D.B., L.R. and C.B. designed the questionnaire with input from all other authors. All authors contributed to developing the questions of the paper. D.B. wrote the first draft of the paper and conducted the analysis, with input from N.B. All authors helped to interpret the data, and review and edit the paper.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15218-2>.

**Correspondence** and requests for materials should be addressed to D.E.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022