

Why Do You Need This? Selective Disclosure of Data Among Citizen Scientists

Anna Rudnicka
 UCL Interaction Centre
 University College London
 London, United Kingdom
 anna.rudnicka.15@ucl.ac.uk

Anna L. Cox
 UCL Interaction Centre
 University College London
 London, United Kingdom
 anna.cox@ucl.ac.uk

Sandy J.J. Gould
 School of Computer Science
 University of Birmingham
 Birmingham, United Kingdom
 s.gould@cs.bham.ac.uk

ABSTRACT

Recent scandals involving data from participatory research have contributed to broader public concern about online privacy. Such concerns might make people more reluctant to participate in research that asks them to volunteer personal data, compromising many researchers' data collection. We tested several motivational messages that encouraged participation in a citizen science project. We measured people's willingness to disclose personal information. While participants were less likely to share sensitive data than neutral data, disclosure behaviour was not affected by attitudes to privacy. Importantly, we found that citizen scientists who were exposed to a motivational message that emphasised 'learning' were more likely to share sensitive information than those presented with other types of motivational cues. Our results suggest that priming individuals with motivational messages can increase their willingness to contribute personal data to a project, even if the request pertains to sensitive information.

KEYWORDS

ACM Classification Keywords: Security and privacy > Human and societal aspects of security and privacy. Author Keywords: citizen science, privacy, disclosure, motivation, dabblers.

ACM Reference format:

Anna Rudnicka, Anna L. Cox, and Sandy J.J. Gould. 2019. Why Do You Need This? Selective Disclosure of Data Among Citizen Scientists. In *Proceedings of CHI Conference on Human Factors in*

Computing Systems Proceedings, Glasgow, Scotland UK, May 4–9, 2019 (CHI 2019), 11 pages.

<https://doi.org/10.1145/3290605.3300622>

1 INTRODUCTION

During the 2016 US presidential election campaign, a group of electioneers had very detailed information about the voters they were trying to lure. They knew so much because they had the profiles of 'at least 50 million (estimated by Facebook themselves to be up to 87 million) global Facebook users' (p.16) [14]. In what is called the Cambridge Analytica scandal, electioneers used these profiles to accurately target election advertising.

The data at the core of the scandal was procured via a personality self-test app promoted through Facebook. The app, 'thisisyourdigitallife', was created by a University of Cambridge academic and promised to give participants some insight into their personality. Approximately 320,000 Facebook users took the test. Critically, participants gave the app permission to access profiles of their Facebook friends. Thus, 320,000 people curious about their personality yielded at least 50 million valuable Facebook profiles.

Researchers often tell people something about themselves in exchange for their participation in research studies [34], [29]. This kind of quid pro quo is an effective way of recruiting new and curious 'citizen scientists' who help to improve our understanding of the world. As citizen scientists, lay-people gather, disclose, or aid the analysis of data as part of an organised, often academic-led, endeavour to further scientific discovery [31]. On Zooniverse [32], a popular citizen science platform, volunteers often complete benign tasks; they help to classify galaxies, count penguins, or transcribe historical manuscripts. In some projects, though, individuals are required to directly or indirectly disclose personal information. For example, the Bandicoot Sighting Register (hosted on SciStarter, the US alternative to Zooniverse), asks citizen scientists who spot a bandicoot, to take a photograph and submit it, alongside GPS location, to the website. Another project, Mappiness, aimed at establishing links between mood and daily environment, asked citizen scientists to record information about how they felt, while



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2019 Copyright is held by the owner/author(s).
 ACM ISBN 978-1-4503-5970-2/19/05.
<https://doi.org/10.1145/3290605.3300622>

the sensor in their phone tracked their location [21]. With the increasing popularity of online citizen science, previous work has signalled the need for human-computer interaction researchers to study this form of data collection in order to maximise the benefits it offers: participant privacy is highlighted as one of the main areas that could benefit from the skills of HCI researchers [28].

The General Data Protection Regulation, which became binding in the European Union in 2018 [9], places greater responsibility on project organisers to obtain fully informed consent from those contributing data. As the new laws make it necessary to inform participants exactly and explicitly what will happen to their data, this precludes any kind of ‘finessing’ of project descriptions. This could easily make citizen scientists more cautious about contributing, even if the underlying data collected has not changed. These changes mean that project coordinators will likely need to put more care and energy into communicating the benefits of taking part in their projects to prospective participants in order to maintain current recruitment rates.

We investigated the impact of reinforcing the benefits of citizen science participation (operationalised by presenting a motivational message at the start of the project) on data disclosure among citizen scientists. We aimed to clarify whether the way in which citizen scientists are encouraged to take part in a project, has an impact on how much personal information they disclose.

The results of our study demonstrate a practical way to encourage the necessary disclosure of sensitive personal data from citizen scientists in a way that promotes transparency and complies with new, stricter privacy laws. We hope that our work will help researchers to design future citizen science projects in a way that is better aligned with people’s attitudes and preferences in relation to data privacy and data disclosure. Our work is intended as a building block towards a more empirically-driven design of online citizen science projects, where data disclosure is approached as part of a transparent, fair and mutually beneficial collaboration.

Moreover, we contribute to research in the area of human-computer interaction by providing an ethical framework for encouraging online participants to disclose sensitive data. We propose that findings from studies such as ours must be utilised to design projects that overtly inform participants about any motivational measures used to enhance disclosure. As will be highlighted later in this paper, we make a clear distinction between coercing participants to disclose data and learning about people’s preferences in order to provide more attractive rewards to participants who decide to ‘donate’ their data to unpaid research projects. While the first practice cannot build lasting trust and cooperation in citizen science communities, the latter can enable project coordinators to remain effective at data collection at a time

when transparency in data collection practices becomes increasingly important. The study presented in this paper sets out to answer the question: ‘can the use of motivational cues increase citizen scientists’ willingness to share sensitive data?’. The study is conducted in a citizen science context, with a sample recruited from the general population rather than, as is the case with most prior research investigating the behaviour of citizen scientists, with small samples of committed long-time contributors. The current study sets out to test three hypotheses:

H1: Privacy fundamentalists will be less willing to disclose information than those categorised as privacy pragmatists and privacy unconcerned.

H2: Participants will be more likely to disclose information when responding to neutral questions than when responding to sensitive questions.

H3: Motivational messages pertaining to learning opportunities, social proof and contribution will result in higher levels of disclosure than messages pertaining to altruism. That is, we expect there will be an interaction between ‘motivation type’ and the impact of ‘item sensitivity’ on disclosure behaviour.

2 RELATED WORK

Citizen Science and Data Privacy

Citizen science has been portrayed as a tool that can simultaneously empower scientists to analyse larger datasets as well as improving scientific literacy among the general public [2]. It is also increasingly being used in HCI and social science, and facilitates collection of scientific data [17]. Participants are often recruited in large numbers through social media or citizen science platforms. They consent to taking part, motivated by a willingness to learn, contribute or connect [8]. In contrast to traditional research, citizen scientists typically do not receive remuneration for taking part in citizen science projects [15], however this does not appear to negatively impact the quality of data. Wiseman et al. demonstrated the feasibility and effectiveness of HCI research with participation of citizen scientists. They found that data provided in the absence of monetary reward, and contributed for altruistic reasons, can equal or indeed exceed the quality of data generated by paid participants [34]. Their results suggest that citizen science can provide viable solutions for large scale HCI experiments that could be otherwise prohibitively expensive to run.

Despite the value of citizen science projects, it is important to note that the design of ethical citizen science requires careful consideration of associated risks that participation might pose to individuals. Citizen scientists are, by default, placed in an uncertain situation. They contribute to activities typically undertaken by trained research staff and are often

simultaneously tasked with roles similar to those of research participants. Citizen scientists do not, however, benefit from nearly the same level of training, or protective measures as professional scientists and traditional lab-based research participants.

These unpaid volunteers often contribute information that they gather with their mobile devices (such as in the case of monitoring noise levels, [19] or else information about themselves, for example when taking part in psychological or health-related citizen science [6]. Some examples of the highly-sensitive data regularly disclosed by citizen scientists are: location (this can relate to personal location at a given point in time, location of work or home, frequency of visiting a particular location), medical data, and information about their work and educational status. These examples of sensitive data are often collected for as meta data, for example a time- and location-stamped photograph of a rare species of bird. Such data, if misused, might yield a way to learn about the routines of a citizen scientist - where they go, when and how long they stay there, or which type of mobile device they can afford. This kind of leakage becomes more concerning when people, for instance, use their phone to constantly monitor noise levels in their neighbourhood. Sharing this recording with researchers could reveal movements and whereabouts over a lengthy period of time. Similarly, an individual completing a psychological survey can easily disclose information that, if misappropriated, could compromise their privacy and security.

The data collected by citizen science projects is not always completely transparent to contributors. Seeking fully informed consent to use or reuse data is not always seen as a priority by citizen science coordinators: only 10 projects (8%) out of the 118 surveyed asked their participants to sign an explicit informed consent form [30]. This could be the result of shying away from discussing any issues that might lower participation. Indeed, it has been demonstrated that making privacy issues salient immediately before disclosure, even if the matter is raised in a positive tone, can cause participants to become more cautious about disclosure [23]. This approach however, presents two problems. Firstly, from an ethical standpoint, whilst the omission of discussion around privacy concerns might be less damaging to long term participants who have a well-rounded understanding of the nature and complexities of information disclosure in a citizen science context, individuals who have had very little experience with citizen science or those who prefer to only 'dabble', (i.e. take part occasionally), may be more at risk of disclosing information about themselves without fully understanding the potential implications or without taking privacy measures (e.g. not sharing with other citizen scientists that the location stamped photograph was taken at their home). It

could be argued that these 'dabblers' may not be fully capable of making an informed decision about disclosing or withholding information in a citizen science project, when project coordinators decide not to discuss privacy issues with participants. Secondly, from a legal standpoint, the newly introduced General Data Protection Regulation [9] makes it impossible to gloss over the specifics of how data will be used or re-used once collected.

It seems likely that the atmosphere of public concern about privacy could put at risk the sustainability of projects that rely on personal data. According to the Ofcom reports from 2015, 2016 and 2017 [25], [26],[27], a trend for growing privacy concerns among internet users can also be observed in the United Kingdom. It is possible that rising internet literacy is offering individuals more insights into the potential dangers that may co-exist with the benefits of smartphone applications and, in general, mobile computing. This growing awareness of online privacy issues might make it more difficult to attract the wider public to take part in disclosure-heavy citizen science projects.

Disclosure of data in citizen science contexts is a particularly under-researched area [4]. The goal of ensuring participants' privacy is noted in citizen science literature [17]. Nevertheless, only one study so far has empirically investigated the issue of privacy with citizen science volunteers: Bowser et al. [3] conducted focus groups and semi-structured interviews with fourteen experienced citizen scientists and thirteen citizen science coordinators. These authors found that the experienced citizen scientists who took part in their study prioritised openness and sharing over privacy. Some participants' responses suggested that this could be a result of a strong motivation to contribute to the project. Discussing these results, Bowser et al. underscored that information disclosure in citizen science differs from other types of information disclosure specifically because of the collaborative, social context of citizen science. However, their participants, described by the authors as 'particularly engaged volunteers familiar with the culture and norms of citizen science' (p. 2127) differ significantly from the majority of citizen scientists, who occasionally 'dabble' in projects rather than dedicating large quantities of time to them [10]. These 'dabblers' who participate in a casual and irregular manner have also been shown to be more motivated by outcomes of a project rather than by the intrinsic pleasure of participation [10]. It follows, then, that defining and emphasising the concrete benefits of participation is especially important when recruiting and retaining 'dabblers', and also possibly a deciding factor when encouraging dabblers to contribute personal information to citizen science projects. It is likely that dabblers, not being attached to or knowledgeable about any particular project, will show more scrutiny when making decisions about whether to disclose or withhold their

data. It is important to focus on the behaviour of dabblers as some citizen science projects (for example those in which the participants are also the subject of study, as can be the case with citizen psych-science, [15] cannot rely on a handful of top contributors completing most of the work. The successful future of many citizen science projects may well depend on the willingness of dabblers to share personal information about themselves.

There are many factors that could be of interest when investigating what causes citizen scientists to disclose or withhold data. Our choice to study the role of motivational cues and existing privacy concerns will enable us to investigate the degree to which participants can be ‘swayed’ to contribute data. Moreover, we aim to compare the disclosure behaviour of ‘dabblers’ to the disclosure behaviour of top contributors studied by Bowser et al. [3] who found that the strong motivation to participate in citizen science observed among top contributors can override the privacy concerns they might otherwise hold. We decided to examine the impact of these privacy concerns and motivational cues against two types of data: neutral and sensitive. This distinction, between neutral data disclosure and sensitive data disclosure, is a reflection of what can happen in a natural citizen science context. For example, a bird watching project [33] may ask us to provide a description of birds we saw during a walk (neutral information), but also enquire about when and where we took that walk (sensitive information that, in case of a data breach, could be misused).

Participant Attitudes

The seminal work of Westin proposes that individuals can be classified as privacy unconcerned (those who do not mind disclosing information about themselves), privacy pragmatists (those who disclose information strategically) and privacy fundamentalists (those who prefer to avoid disclosing information about themselves), [11]. In a recent study, participants were required to complete a credit card application that asked for sensitive information. Individuals who were classed as either privacy pragmatists or unconcerned according to the Westin Privacy Scale, were 5.6 times more likely to submit the form, when compared to privacy fundamentalists [22].

Nevertheless, privacy research has shown that ability to predict disclosure behaviour from the level of privacy concern is inconsistent, a phenomenon referred to as the ‘privacy paradox’ [18]. It has been observed that even when individuals state that they are concerned about their privacy, it is still relatively easy to convince them to part with their data. For example, in a field experiment conducted by Beresford et al. [1], participants were asked to buy a DVD from one of two stores, both of which asked for some personal information: the income and date of birth in the case of the first store,

and favourite colour and year of birth in the second store. Despite the first store requesting far more sensitive information, when the price of DVD was the same, participants bought from both stores equally often. When the price of the store asking for more sensitive information was lowered only by one euro, the great majority of participants chose that store. Notably, post-experiment testing showed that 95% of individuals taking part in this study were interested in protecting their data.

In light of the conflicting evidence described above, we examine whether attitudes to privacy influence disclosure of information in a citizen science context. As our current study gives the participants an opportunity to leave questions unanswered and still continue with participation, we predict that the disclosure-shy privacy fundamentalists will be retained within the study, but present with lower disclosure on items that pertain to sensitive topics or enquire about data of third persons.

Data Sensitivity

An important variable for studying disclosure behaviour is the sensitivity of information requests. For example, asking about a person’s favourite colour is a neutral request when compared to asking them about their address or date of birth. In a credit card application context, Malheiros et al. [22] found that participants were much less likely to cooperate with an information disclosure request when questions pertained to sensitive information such as health records or friends’ email addresses, as opposed to more neutral questions pertaining to participants’ demographics. It may be important for researchers designing citizen science projects, to recognise the difference between neutral and sensitive information requests. It is possible that, with privacy awareness growing in the wider population [27], requests to disclose sensitive personal data may need to be coupled with additional design features such as motivational cues or pre-testing participants for privacy attitudes. We expect that the impact of data sensitivity on disclosure demonstrated by Malheiros et al. [22] will also feature in a citizen science context. Therefore, we examine whether the sensitivity of a data item will be correlated with the proportion of participants responding to it.

Motivation to Participate

The use of monetary rewards to help convince individuals to part with their personal data is a common focus for privacy studies [35]. It has been proposed that most disclosure decisions involve a privacy calculus scenario, where individuals weigh the perceived cost of disclosure against the benefits of disclosing information about themselves [16]. While this calculation is easy to observe and manipulate in the private sector (for example a certain amount of money or a free

service can be offered in exchange for personal data), the voluntary context of citizen science prompts the study of how non-monetary rewards may impact data disclosure decisions. Discussing the results of their study, Bowser et al. [3] brought to attention the link between what citizen scientists perceived as rewards of participation, and the disclosure decisions they made while taking part in citizen science projects. Although the role of motivation in citizen science has been the focus of many investigations, the study conducted by Bowser et al. is the first to empirically establish a link between the motivation of a citizen scientist to participate in a project and their willingness to part with personal data.

Studies of participant recruitment have shown that priming potential citizen scientists with motivational messages can have an impact on their behaviour. In a large study (n=36,513) conducted by Lee and colleagues [20], potential participants were sent emails with one of four randomly assigned messages in the subject line. The four variations of the message all focused on trying to recruit participants for the Gravity Spy project, however each message focused on a different motivation to take part: 1. Learning science ('Extend your knowledge in astrophysics by participating in Gravity Spy!'); 2. Social Proof ('Join your fellow citizen scientists in classifying problematic noise in the search for gravitational waves!... Many citizen scientists are already participating in the project!'); 3. Contribution to science ('You can contribute to science by classifying problematic noise in the search for gravitational waves!'); 4. Altruism ('Astrophysicists need your help to classify problematic noise in the search for gravitational waves!'). Lee et al. found that messages emphasising learning, social proof and contribution were more effective in attracting participants than those alluding to altruism. Moreover, Diner et al.[7] have demonstrated the importance of reinforcing the social context of citizen science; they found that citizen scientists who were presented with a message about the contributions of a high-performing peer or group would increase their own contribution level. It may be important for researchers designing citizen science projects to consider how to motivate volunteers to participate. In our study we explore whether 'motivation type' impacts disclosure behaviour.

3 METHOD

Participants

Participants were recruited online via Twitter and through word of mouth. While recruiting on Twitter, we employed the use of paid twitter ads that promoted our tweets across participants who had used the hashtag #citizenscience. This enabled us to reach a wider population of individuals open to participating in citizen science, and avoid the recruitment of a

sample made up mostly of individuals already involved in academic research. While a small subset of participants was recruited through word of mouth and a link to the website was emailed to them, most participants were recruited via social media, specifically, on Twitter. Tweets encouraging participation were posted, for example: 'Did you get enough ZzZzZzs last night? Tell us about your sleep habits in this awesome #citizenscience survey! <http://www.sleepmapping.net>'; 'It's my party and I'll sleep if I want to... Take part in our #citizenscience project about SLEEP <https://sleepmapping.net>'; or 'How many ZzZzZzs are you getting? Tell us! [sleepmapping.net](http://www.sleepmapping.net) #citizenscience'.

Participation was open to all individuals over the age of 18, irrespective of their location of residence. Of the 566 individuals who followed the link to the survey, 331 decided to take part in the study. Participants who dropped out throughout the study (n=149) were excluded. Solely the data of those individuals who completed the final question of the survey were selected for final analysis. The resultant sample consisted of N=182 participants (128 female, 53 male, 1 undisclosed), with a mean age of 38 years.

Materials

Materials included a website, www.sleepmapping.net, created for this project and a survey hosted online, via a survey platform www.qualtrics.com. The website contained a message that said 'Welcome. Take part in the Sleep Mapping survey!' and a link to the Qualtrics-based survey.

The survey consisted of the following elements: a motivational message (1 out of 4, randomly assigned), a Participant Information Sheet, sixteen consent questions compliant with the General Data Protection Regulation (2016), a question enquiring about interest in future studies, Demographics-related questions (age, gender, Internet use, previous citizen science participation), 19 Neutral Item Questions, 14 Sensitive Item Questions, and the Westin Privacy Scale.

The motivational messages were adapted from a study conducted by Lee et al. [20], and were as follows: 1. 'Extend your knowledge of health psychology by participating in the Sleep Mapping survey!' ('Learning' motivation); 2. 'Join your fellow citizen scientists in establishing connections between stress and sleep! Many citizen scientists are already participating in the project.' ('Social Proof' motivation); 3. 'You can contribute to science by answering questions about sources of stress in your life and quality of sleep!' ('Contribution' motivation); and 4. 'Health psychology needs your help to connect sources of stress to patterns of sleep behaviour' ('Altruism' motivation).

The Morningness-Eveningness Questionnaire [13], served as Neutral Questions in this study. The survey was presented in its original form, with the exception of questions 1,2,10,17 and 18. In these five questions the mode of response was

changed from marking a place across a bar, to a multiple-choice format. This was done to better suit the online environment of this experiment.

The 14 Sensitive Questions were adapted from Malheiros et al. [22]. As citizen science relies on voluntary contributions, in contrast to Malheiros et al, we did not make it compulsory to answer each question. This survey therefore made it possible to skip certain questions and still progress to the next item. Moreover, in questions that asked participants to input information about third persons that they might not have or be confused about, we added an option of 'I prefer not to answer' to distinguish between unwillingness to answer and lack of sufficient information (such as email addresses of friends or the address of previous employer).

The final part of the survey was the 3-question Westin Privacy Scale [11] which required the participants to judge the truthfulness of the following three statements alongside a 4-point Likert type scale, (1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, and 4 = strongly agree): 1. 'Consumers have lost all control over how personal information is collected and used by companies'; 2. 'Most businesses handle the personal information they collect about consumers in a proper and confidential way'; and 3. 'Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.' This scale enables classification of participants as either privacy fundamentalists who are defensive about sharing personal data (agreed with the first statement and disagreed with the second and third statements), privacy unconcerned who do not have major concerns about parting with their personal data (disagreed with the first statement and agreed with the second and third statements). All other patterns of responses to these statements classify participants as 'privacy pragmatists', that is individuals who weigh the pros and cons of sharing their data and are willing to disclose information when they feel it is beneficial and justified.

Procedure

The link advertised on Twitter led participants to the dedicated website created for this project under domain name www.sleepmapping.net. There, a welcoming message was followed by a link to the survey, hosted on the Qualtrics platform. After being redirected to the Qualtrics website but before starting the survey, participants were presented with one of the four randomly assigned 'motivational messages', adapted from Lee et al. [20]. After consenting to take part, participants filled the 19-item Morningness-Eveningness Questionnaire [13], which served as the Neutral Questions in this study. Secondly, participants were asked to answer 14 questions, adapted from the study conducted by Malheiros et al. [22], which served as Sensitive Items. Finally, participants were presented with the Westin's Privacy Scale [11] and

asked to make three judgements about the degree to which they agree with three statements about privacy. Following the survey, participants received a debriefing message.

It is important to note that this study used deception. To emulate the authentic context of citizen science, we created a project called 'Sleep Mapping'. Participants were told that they are taking part in a study, which investigated the connection between sources of stress and sleep patterns. This element of deception was used for several reasons. Firstly, as noted by Bowser et al. [3], coordinators of citizen science projects are not keen to 'share' their participants, for fear that it could shift the focus away from the original purpose of their initiatives, and it would therefore have been difficult to recruit participants from an existing project. Secondly, we were keen to conduct current research in the context of a new project, so that the recruited sample would be more representative of 'dabblers', as opposed to a select sample of committed participants. Finally, it should be noted that in the current study, the participant information sheet presented before the consent form, informed individuals about the possibility of deception within the study; the debriefing form seen by participants at the end of the study also fully explained why deception was required.

4 RESULTS

Demographics

The link to the online 'Sleep Mapping' survey was followed by 566 individuals, of which 331 decided to take part in the study. Of those, participants who dropped out throughout the survey (defined here as not having reached the final survey question) were excluded ($n=149$). Although individuals who followed the link to the survey were randomly assigned to four Motivation groups in equal numbers, due to the pattern of attrition in this sample (participants dropping out throughout the survey), the final sample ($n=182$) included slightly uneven numbers of participants in the 4 motivational groups: Learning ($n=42$), Social Proof ($n=40$), Contribution ($n=50$), and Altruism ($n=50$).

The final $n=182$ sample consisted of 128 females, 53 males and 1 participant who did not disclose their gender. Similarly, data about age of participants were available for $n=181$; the ages of participants ranged from 19 to 71 years old, with a mean of 38 years ($SD=11.5$). Likewise, data about Internet use were available for 181 participants, with 115 participants stating that they 'use the Internet all the time', 61 who 'use the Internet several times per day' and 5 who 'use the Internet most days'. None of the participants chose the option indicating the use of the Internet 'less than once a week' or 'less than once a month', indicating that this sample was likely to rely on the Internet for many of their daily tasks.

None of these three variables (gender, age, Internet use) had an impact on disclosure of information.

With regards to previous citizen science experience, more than half of the 182 participants had never taken part in a citizen science project before ($n=115$), while 46 had ($n=20$ took part once before, and $n=26$ had taken part several times); 21 participants were ‘not sure’ about whether they had participated previously or not. We found no significant differences on disclosure between these groups. Moreover, it should be specified that, in the context of this study, all participants are technically dabblers as this particular project only offers the opportunity to participate once (i.e. to dabble with the Sleep Mapping project).

Participant Attrition

In this study, the sharpest decline in participant numbers could be observed at the first stage of recruitment: while 566 participants clicked the link to the survey and were presented with one of the four randomly assigned motivational messages, only 332 proceeded to view the Participant Information Sheet and answer the main consent question. 331 participants declared that they wished to proceed with the study (they said ‘yes’ to the main consent question). Attrition over the following 17 consent questions (which were each presented on a separate page requiring 2 clicks (‘yes’ and ‘next’) to proceed) resulted in a sample of $n=240$ at the beginning of Demographics items.

Neutral items were all presented on one page. Twelve participants dropped out across these items. A sharp decline in participant numbers was then observed at the point of the first sensitive question: out of 228 participants who answered the last Neutral Item, only 205 answered the first Sensitive Item. This 10.1 % relative decline at this point in the survey suggests that participants paid attention to the level of sensitivity of the questions they were being asked.

Nevertheless, only 10 participants dropped out over the course of Sensitive Items, a comparable number to the 8 participants who dropped out throughout the Neutral Items part of the survey. This suggests a greater role of the point of change from neutral to sensitive questions in the survey, than the cumulative impact of sensitive questions, for participant attrition.

A final sharp decline can be observed when progressing from the last sensitive questionnaire, to the Westin privacy scale: here, 13 participants drop out before the Westin Scale. In total, a sample of $n=182$ completes the last item of the survey.

Hypothesis 1

Hypothesis 1 assumed that ‘privacy fundamentalists’ would be less likely to disclose information than those categorised as privacy pragmatists and privacy unconcerned. Based on

their answers on the 3-item Westin Privacy Scale, participants were classified as privacy fundamentalists ($n=71$), privacy pragmatists ($n=100$), and privacy unconcerned ($n=11$). Across this classification, privacy pragmatists showed slightly higher overall disclosure ($m=24.58$, $SD=2.07$) than privacy fundamentalists ($m=24.18$, $SD=2.29$) and privacy unconcerned ($m=23.45$, $SD=1.44$). However, a one-way between-subjects Analysis of Variance showed that the impact of Westin group on overall disclosure was not statistically significant ($F(2,179) = 1.431$, $p = .242$, partial $\eta^2 = .016$). Hypothesis 1 was rejected.

Hypothesis 2

We predicted that participants would be more likely to disclose information when responding to neutral questions than when responding to sensitive questions. Disclosure on the neutral survey items (Circadian Rhythm Questionnaire) that consisted of 19 questions was very high, ranging from disclosure on 17 items to disclosure on all of the 19 items (mean= 18.98 , $SD= 0.16$). This resulted in a mean 99.89% proportion of disclosure on neutral survey items. Disclosure on the 14 sensitive items was low. It ranged from 3 disclosed items to 12 disclosed items (mean= 5.37 , $SD=2.12$). This resulted in a mean 38.36% of disclosure on sensitive survey items. A paired t-test showed that the difference between conditions (sensitive vs neutral) was significant ($t = 54.901$, $df= 181$, $p=.000$, 2-tailed). Hypothesis 2 was accepted.

It should also be noted that, in certain instances, participants included comments on questions even when disclosure of information did not happen. For example, several participants, when asked questions about third party data noted that they ‘can provide if needed but must ask for their permission first’. Another participant noted that they would be able to provide the information however they didn’t feel that they had been warned about the possible collection of this type of data (‘Yes, but you certainly didn’t ask me in advance if I was OK to share that kind of information’). Another participant left a comment saying, ‘why do you need this?’. Such quotes suggest that participants maintain an alert and analytical attitude towards the types of disclosure requests they are faced with.

Hypothesis 3

Hypothesis 3 predicted an interaction effect between the impact of ‘motivational message’ and the impact of ‘item sensitivity’ on level of disclosure. Table 1 displays the mean disclosure on neutral and sensitive items across the four motivational groups. A mixed 4×2 Analysis of Variance showed that there was a statistically significant interaction effect of item sensitivity and motivation group interaction was significant: $F(3,178) = 3.717$, $p = .013$, partial $\eta^2 = .059$. Hypothesis 3 was accepted.

Table 1: Mean Disclosure Across Motivational Groups

	N=42 LEARNING	N=40 SOCIAL PROOF	N=50 CONTRIBUTION	N=50 ALTRUISM
Neutral items	19 (0.00)	19 (0.00)	18.94 (0.31)	19 (0.00)
Neutral Items %	100% (0.00%)	100% (0.00%)	99.68% (1.65%)	100% (0.00%)
Sensitive Items	6.31 (2.94)	5.00 (1.47)	5.02 (1.89)	5.26 (1.76)
Sensitive Items %	45.07% (21.02%)	35.71% (10.48%)	35.86% (13.49%)	37.57% (12.56%)

Post hoc comparisons of motivational groups were conducted to determine the relative importance of each of the 4 motivational cues for influencing disclosure of Sensitive Items. We found that the Sensitive Item disclosure in the Learning group was significantly higher than in the Social Proof ($t = 2.530$, $df = 80$, $p = .013$, 2-tailed), Contribution ($t = 2.539$, $df = 90$, $p = .013$, 2-tailed), and Altruism ($t = 2.113$, $df = 90$, $p = .037$, 2-tailed) messages. Moreover, the differences between Social Proof and Contribution ($t = -.055$, $df = 88$, $p = .956$, 2-tailed), Social Proof and Altruism ($t = -.749$, $df = 88$, $p = .456$, 2-tailed), and Altruism and Contribution ($t = -.657$, $df = 98$, $p = .513$, 2-tailed), were not found to be statistically significant. These results suggest that it was the Learning message solely that was responsible for the differences between Sensitive Item disclosure across the Motivational Group variable.

5 DISCUSSION

This study investigated the degree to which motivational messages, privacy attitudes and item sensitivity influence data disclosure behaviour in a citizen science setting.

Participant Attrition

Following the initial dropout that can be expected in Internet-based survey research [12], where 234 participants dropped out before viewing the Information Sheet and a further 92 dropped out over the consent form, sample size remained relatively stable throughout the survey.

While filling the survey, comparable numbers of participants dropped out during the Neutral Items (8 participants over 19 questions) and Sensitive Items (10 participants over 14 questions). This suggests that once committed to answering a block of questions, citizen scientists do not leave the survey even when asked about sensitive data. This conclusion is further supported by the fact the greatest post-consent decline (23 participants) happened just before the first Sensitive Item, suggesting a greater role of the change in the sensitive of items and a smaller role of accumulative impact of many sensitive questions. It is nevertheless clear that presentation of sensitive questions appears to detract individuals from finishing the survey.

Future research might explore whether presenting neutral and sensitive items together as one block would result in a different pattern of attrition. Alternatively, reminding participants that it is important to remain within the survey, even if some questions are skipped and unanswered, could help clarify the relationship between privacy-related attrition and other reasons that may have caused dropping out such as boredom.

Hypothesis 1

Contrary to our prediction, we found no evidence to suggest that ‘privacy fundamentalists’, as classified by answers to the 3-item Westin Privacy Scale [11], were less likely to disclose sensitive information than ‘privacy pragmatists’ or ‘privacy unconcerned’. It has been argued that personal attitudes towards privacy can be overridden by perceived rewards to be gained in exchange for disclosing data [24]. It is possible that individuals in this study perceived the rewards of citizen science participation to be greater than the ‘cost’ of disclosing personal data, even if they did have concerns about privacy. This is in line with the results of Bowser et al. [3] who found that participants decided to share information even when they had some privacy concerns, because their concerns were overridden by the specific motivations they had to take part in citizen science in the first place.

It should be noted that at this point we are not able to directly compare the results demonstrated by Malheiros et al. [22] from whom we adapted the sensitive item questions. Malheiros et al. found that privacy fundamentalists were less likely to submit the form in their study. In the current study, the Westin scale was administered at the very end of the survey (and therefore only to participants who have not dropped out prior to this point), thus we were not able to observe the possible impact of the Westin scale classification on attrition of participants. This was a methodological consideration necessary to avoid bias and maintain the authenticity of the citizen science project. In the future however, it may be useful to contact those participants who dropped out and ask them to complete the Westin scale. This could help clarify whether privacy concern has any impact on participant attrition. Overall, however, our results provide evidence to

support the existence of a discrepancy between people's self-reported privacy concern and their disclosure behaviour in realistic scenarios [24].

Hypothesis 2

Replicating the results demonstrated by Malheiros and colleagues (2012), we found that participants were more likely to disclose data when faced with neutral questions than when answering questions about sensitive issues. Despite having limited knowledge about the study's purpose, participants made their own judgements about whether a question was relevant to the study's aim or not.

Similarly to Malheiros et al. participants, our sample was particularly resistant to answering questions about third parties. From an ethical standpoint, this is a very positive finding, suggesting that individuals are likely to respect the privacy of third parties, and that this is true across different contexts. The comments left by citizen scientists suggest that they are far from being passive research participants - they maintain an alertness regarding the types of disclosure requests they face. It is likely that citizen scientists are willing to cooperate with researchers even when sensitive data are requested, however they require full explanation of why the researcher needs a particular type of data. This is in line with the findings of Malheiros et al. [22], who demonstrated that participants were less likely to share information if they perceived the questions to be irrelevant to the overarching purpose of the form they were filling. However, it is also possible that participants were influenced by the name of the study, 'Sleep Mapping', to be more accepting of sleep-related questions, and that if the necessity of investigating stress sources was emphasised more saliently (for example by calling the project a Sleep-Stress Study), then some of the Sensitive Item questions could have appeared less invasive and more relevant. This hypothesis could be clarified in future research by expanding the range of both the neutral questions and the sensitive questions, as well as by conducting post-survey interviews with participants.

Hypothesis 3

We also found an interaction effect between the type of motivational priming and item sensitivity: participants primed by the 'Learning' motivational message, were more likely to disclose sensitive data, than participants primed by other motivational messages. Motivational message did not have an impact on disclosure across Neutral Questions. Firstly, it should be noted that due to the ceiling effects in the Neutral Question condition, we do not know whether participants simply do not require additional motivation to disclose information when faced with non-sensitive questions, or whether this could be due to the close connection between the Neutral items and the name of the project (as briefly discussed in

relation to Hypothesis 1). Overall, however, it appears that presenting a motivational message can lead to a participant disclosing more personal information, even when the data requests pertain to sensitive personal information. This is in line with both the link between motivation and disclosure found by Bowser et al. [3] as well as the theoretical link between the importance of clarity in communication about the aims of collecting data and the ability of data collectors to successfully gather personal information about data subjects [5].

From a practical standpoint, our findings suggest that emphasising certain types of motivations can enhance the volume of disclosed data and therefore support the primary goals of projects that are focused on gathering personal data from citizen scientists. It should however be clarified whether 'Learning' is the primary type of motivation that will encourage disclosure of information or whether different samples may demonstrate variable sensitivity to different types of motivational cues. Future research should explore the stability of the motivation-disclosure relationship across various participant-specific variables such as personality traits, cognitive style, or mode of citizen science participation (casual vs committed).

Despite the fact that our findings reached statistical significance, it should be noted that the differences in Sensitive Item disclosure across motivational groups were relatively small (45.07% of sensitive-item disclosure in participants primed by Learning, compared to 35.71% for participants primed by Social Proof, and 35.86% and 37.57% for those who were presented with Contribution and Altruism messages, respectively). In order to identify useful design implementations that could help increase data disclosure in citizen science projects, more research needs to be conducted, both exploring efficacy of motivational cues in larger samples, as well as investigating participant sensitivity to those cues.

Finally, it should be noted that this study was conducted with a sample that consisted primarily of 'dabblers' (casual contributors). Future research should address the question of to what extent motivational cues can encourage disclosure of data across different samples, such as committed citizen scientists or students who participate in online research projects in exchange for course credit.

6 LIMITATIONS

This study explored only a limited range of sensitive and neutral questions and including a wider range of questions in future research could help distinguish between withholding of information due to perceived irrelevance of question to the topic of study vs due to the level of sensitivity. Moreover, we do not have much information on why some participants dropped out - whether they were bored or unmotivated or whether they found the survey too intrusive.

Motivational cues were only presented to participants as a very brief pre-participation message; in the future, it may be informative to ask participants to make a conscious choice about the source of motivation they wish to prioritise while participating in citizen science - in order to ensure deeper cognitive processing of the motivational aspect of the study. It should be explored whether such a study can be conducted effectively and rigorously without the use of deception. Future research must explore, both theoretically and empirically, how to best implement motivational priming in a way that remains within the boundaries of fair exchange. In other words, it would not be ethically sound to prime unsuspecting individuals to disclose more information than they would have wanted to. Rather, the purpose of studying motivational priming should be to support the principle of fair exchange in citizen science and to learn what type of rewards are most favoured by citizen scientists, in the context of projects that require a relatively high level of disclosure of potentially sensitive personal information.

We created our own project ‘sleep-mapping’ for the purposes of this research which required us to deceive our participants about the true purposes of our intentions. An alternative solution could have been to co-create a new project in collaboration with researchers interested in running a real citizen science initiative and collect privacy related data at the same time as allowing participants to contribute data to citizen science. This would have allowed us to minimise the amount of deception. Nevertheless, in the context of the current study, we saw this option as not logistically feasible. In the future, the citizen science research community will benefit from more discussion on how to most productively and ethically conduct research about citizen science and to what extent deception is acceptable.

7 CONCLUSION

People frequently disclose information about themselves in exchange for something they perceive to be valuable. In this study, we demonstrated the existence of selective disclosure of personal data in a citizen science context. Participants were more likely to disclose personal data when primed by a motivational message that emphasised ‘Learning’ opportunities than by other messages. To our knowledge, this is the first study to experimentally investigate antecedents of disclosing or withholding personal information by citizen scientists. Secondly, these results suggest that the degree to which participants disclose information is not affected by their privacy attitudes. This is in line with the privacy paradox phenomenon described as a marked discrepancy between what individuals think about privacy and how they behave in data disclosure situations. It is likely that the simple design feature of presenting brief motivational messages

at the start of citizen science projects may be used to encourage disclosure of scientifically valuable personal data, without having to be circumspect about the data that is being collected and the ways in which it will be used.

8 ACKNOWLEDGMENTS

We would like to thank our participants for contributing their time and effort to our research. We also wish to thank four anonymous reviewers for their insightful suggestions as well as Dr Duncan Brumby for his comments on an earlier draft of this paper.

REFERENCES

- [1] Alastair R. Beresford, Dorothea Kübler, and Soren Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 2012.
- [2] Rick Bonney, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, 2009.
- [3] Anne Bowser, Katie Shilton, Jennifer Preece, and Elizabeth Warrick. Accounting for privacy in citizen science: Ethical research in a context of openness Volunteer Motivation. *Proceedings of CSCW 2017*, 2017.
- [4] Anne Bowser, Andrea Wiggins, Lea Shanley, Jennifer Preece, and Sandra Henderson. Sharing data while protecting privacy in citizen science. *interactions*, 2014.
- [5] Matthew Crain. The limits of transparency: Data brokers and commodification. *New Media and Society*, 2018.
- [6] Lea Den Broeder, Lidwien Lemmens, Serfanim Uysal, Karin Kauw, Jitske Wekenborg, Michaela Schönenberger, Simone Klooster-Kwakkelstein, Mieke Schoenmakers, Willie Scharwächter, Annemarije Van de Weerd, Samira El Baouchi, Albertine Jantine Schuit, and Annemarie Wagemakers. Public Health Citizen Science; Perceived Impacts on Citizen Scientists: A Case Study in a Low-Income Neighbourhood in the Netherlands. *Citizen Science: Theory and Practice*, 2(1):1–17, 2017.
- [7] David Diner, Shinnosuke Nakayama, Oded Nov, and Maurizio Porfiri. Social signals as design interventions for enhancing citizen science contributions. *Information, Communication & Society*, 21(4):594–611, 2018.
- [8] Margret C. Domroese and Elizabeth A. Johnson. Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation*, 208:40–47, 4 2017.
- [9] European Union. Regulation 2016/679 of the European parliament and the Council of the European Union. *Official Journal of the European Communities*, 2016.
- [10] Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L. Cox. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 2014.
- [11] Harris and Associates Inc. and Alan Westin. E-commerce and privacy: What net users want. Technical report, Privacy and American Business and Pricewaterhouse Coopers LLP, 1998.
- [12] Michael Hoerger. Participant Dropout as a Function of Survey Length in Internet-Mediated University Studies: Implications for Study Design and Voluntary Participation in Psychological Research. *Cyberpsychology, Behavior, and Social Networking*, 2010.
- [13] J. Horne and O. Ostberg. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Internacional*

- Journal of Chronobiology*, 4(2):97–110, 1976.
- [14] Information Commissioner’s Office. Investigation into the use of data analytics in political campaigns. Investigation update. 2018.
- [15] Charlene Jennett, Dominic J. Furniss, Ioanna Iacovides, Sarah Wiseman, Sandy J. J. Gould, and Anna L. Cox. Exploring Citizen Psych-Science and the Motivations of Errordiary Volunteers. *Human Computation*, 2014.
- [16] Zhenhui Jiang, Cheng Suang Heng, and Ben C. F. Choi. Privacy concerns and privacy-protective behavior in synchronous online social interactions. *Information Systems Research*, 2013.
- [17] Sunyoung Kim, Jennifer Mankoff, and Eric Paulos. Sensr. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW ’13*, 2013.
- [18] Spyros Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon, 2017.
- [19] Simone Leao, Kok Leong Ong, and Adam Krezel. 2Loud?: Community mapping of exposure to traffic noise with mobile phones. *Environmental Monitoring and Assessment*, 2014.
- [20] Tae Kyoung Lee, Kevin Crowston, Carsten Østerlund, and Grant Miller. Recruiting Messages Matter. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW ’17 Companion*, 2017.
- [21] George MacKerron and Susana Mourato. Happiness is greater in natural environments. *Global Environmental Change*, 23(5):992–1000, 10 2013.
- [22] Miguel Malheiros, Sacha Brostoff, Charlene Jennett, and M. Angela Sasse. Would you sell your mother’s data? personal data disclosure in a simulated credit card application. In *The Economics of Information Security and Privacy*. 2013.
- [23] Helia Marreiros, Mirco Tonin, Michael Vlassopoulos, and M. C. Schraefel. ‘Now that you mention it’: A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior and Organization*, 2017.
- [24] Patricia A. Norberg, Daniel R. Horne, and David A. Horne. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41:100–126, 2007.
- [25] Ofcom. Adults’ Media Use and Attitudes Report 2015. Technical report, 2015.
- [26] Ofcom. Adults’ Media Use and Attitudes Report 2016. Technical report, 2016.
- [27] Ofcom. Adults’ Media Use and Attitudes Report 2017. Technical report, 2017.
- [28] Jennifer Preece. Citizen Science: New Research Challenges for Human-Computer Interaction. *International Journal of Human-Computer Interaction*, 2016.
- [29] Katharina Reinecke, Ann Arbor, and Krzysztof Z. Gajos. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. *CSCW*, 2015.
- [30] Sven Schade and Chrysi Tsinaraki. *Survey Report: Data Management in Citizen Science Projects*. 2016.
- [31] Jonathan Silvertown. A new dawn for citizen science. *Trends in Ecology and Evolution*, 2009.
- [32] Robert Simpson, Kevin R. Page, and David De Roure. Zooniverse: observing the world’s largest citizen science platform. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 1049–1054. ACM, 2014.
- [33] Brian L. Sullivan, Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, and Steve Kelling. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10):2282–2292, 10 2009.
- [34] Sarah Wiseman, Anna L. Cox, Sandy J. J. Gould, and Duncan P. Brumby. Exploring the effects of non-monetary reimbursement for participants in HCI research. *Human Computation*, 4(1):1–24, 2017.
- [35] Allison Woodruff, Vasyl Pihur, Lauren Schmidt, Laura Brandimarte, and Alessandro Acquisti. Would a privacy fundamentalist sell their DNA for \$1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In *USENIX Association Tenth Symposium On Usable Privacy and Security*, volume 1, 2014.