Disclosure of Personal Data in Citizen Science Settings

A thesis submitted to

University College London

For the degree of

Doctor of Philosophy

CANDIDATE

Anna Małgorzata Rudnicka

SUPERVISORS

Professor Anna L. Cox

Dr Sandy J.J. Gould

EXAMINERS

Professor Corina Sas

Dr Aneesha Singh

A dissertation submitted in partial fulfilment of the requirements for the degree of:

Doctor of Philosophy of **University College London**

UCL Interaction Centre, Division of Psychology and Language Sciences,

University College London

DECLARATION

I, Anna Małgorzata Rudnicka, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

An increasing number of research projects rely on data provided by citizen scientists and many of these projects collect personal information about the citizen. We know very little about why a citizen scientist would decide to disclose or withhold their data. With data protection becoming one of the most socially salient issues and the focus of a recent legislative overhaul, it is important to understand the complexities of human behaviour in voluntary disclosure scenarios.

This thesis describes two threads of investigation. Firstly, we present three studies that examined disclosure behaviour in the context of an online sleep survey. Participants in these studies were not part of any existing online community. We demonstrate that, in this context, citizen scientists shared a larger volume of data when prompted by a message that emphasised learning opportunities when compared to messages focused on other types of motivation. Moreover, we show evidence for the distinction between citizen science and online quizzes, demonstrating that the relationship between motivational messages and data disclosure differs across these two types of projects.

Secondly, we report two studies conducted by encouraging members of an online pet owner community to take part in an online citizen science survey. We found that the presentation of affirming motivational messages impacts data disclosure in a different way than asking people about their motivational preferences. Moreover, selfless benefits of citizen science participation were the most popular among the participants, however the least effective at eliciting data disclosure.

The findings presented in this thesis demonstrate that how people are encouraged to take part in citizen science has an impact on the volume of sensitive data that they disclose within these projects. This has implications both for the design of effective citizen science initiatives but also for the ethics of citizen science.

Impact Statement

Growing awareness of the importance of data privacy means that more careful consideration of the fairness of data exchanges is now required. This thesis analyses the disclosure of data by citizen scientists, individuals who volunteer their time and effort for scientific research. We conducted five data disclosure studies, with a total of over a thousand participants (n = 1050). This allowed us to provide practical insights into how the design of citizen science projects can influence the volume of personal data that individuals participating in these projects disclose.

We demonstrate that the promises made to citizen scientists – about the benefits of participation in citizen science projects – can influence their willingness to disclose personal information about themselves and others. Our contribution to the research literature enables researchers to make citizen science projects both fairer and more effective: this thesis shows, both on theoretical and empirical grounds, that it is increasingly important for project coordinators to only communicate about realistic and concrete benefits of participation. This is because communication about the benefits of participation is likely to increase the level of data disclosure. It follows then that any overblown or vague promises made about the benefits of taking part in citizen science can be viewed as a threat to the fairness of the exchange that takes part between citizen scientists and citizen science project coordinators.

By demonstrating what makes people share data, we offer insights that will help citizen science coordinators create projects that are both sustainable and more effectively attract personal data in cases where projects rely on that. We lay a foundation for a more transparent style of communication about the benefits of citizen science, in contexts where citizen science projects require the sharing of sensitive data.

Moreover, we contribute to the field of Human-Computer Interaction (HCI). With an increasing number of HCI studies utilising volunteer participation, and with the growing societal awareness of privacy issues, we provide guidelines for the collection of sensitive data from volunteers. The issues of disclosure in citizen science fit into a greater societal paradigm shift in how data disclosure is regulated and perceived by data subjects. This is relevant to all HCI researchers. Our approach to the ethics of fair exchange between data contributor and data collector will help support the effective yet ethical collection of data by HCI researchers.

Finally, we provide a blueprint for academic research. Namely, we show that it is possible to combine ethics and effectiveness. As academics strive to make a positive impact on public policy, it will become increasingly important to ensure that discussion of ethical issues is not merely theoretical but also data driven. This thesis provides an example of how to approach ethical issues of societal importance from an empirical standpoint.

Acknowledgements

I feel incredibly fortunate to have undertaken my PhD at the UCL Interaction Centre. I am grateful to my supervisors, Anna and Sandy, who believed in me and provided me with excellent guidance and unconditional support. It is a great privilege to be mentored by people whom you trust and admire.

My PhD was generously funded by the Engineering and Physical Sciences Research Council (grant EP/M507970/1).

1050 citizen scientists gave their time and effort. I also got to know a beautiful community of cat owners on Twitter. Many of you took part in my studies and many more became instant friends.

Duncan Brumby examined my 1st Year and Upgrade reports, and Joe Newbold examined my Upgrade report. Your advice was immensely helpful.

UCLIC is a vibrant place to study and I am grateful to all the researchers and students I met and spent time with, I have learnt so much from you and you inspired me to explore new and exciting areas of research. The A-team has kept me on my toes in all the right ways.

Carl, I cannot imagine having done any of this without you, thank you for all the coffee, hundreds of hours of discussions and proofreading, and for letting me constantly move the furniture.

My family – Mum, Dad, Lena and Mikołaj – your continual encouragement to broaden my horizons has helped me become a better researcher and a more confident person.

Leonard Cohen, thank you for the music.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Motivation

Citizen science involves laypersons contributing to research activities, for example by gathering or analysing data. A recent classification by Muki Haklay (Haklay, 2015) identified six distinct citizen science activities, ranging from the relatively low level of involvement required for ‘passive sensing’ and ‘volunteer computing’, through ‘volunteer thinking’ and ‘environmental and ecological observation’ to the more participant-controlled ‘participatory sensing’ and ‘civic/community science’. Citizen science projects play an increasingly important role in scientific research, as a tool for data collection and analysis, as well as a way of involving lay citizens in activities that familiarise them with the scientific process (Bonney et al., 2009). Technological advances such as the increase in availability and affordability of mobile sensing devices and the rise of sophisticated solutions that enable individuals with limited training to capture highly specialised data have led to a series of substantial changes in the practice of citizen science (Preece, 2016).

Participants are sometimes asked to contribute personal information or provide information that, while not focused on citizen scientists as data subjects, may still include sensitive metadata such as their location, time of data collection or IP address. The result is an increasing saliency of issues linked to information disclosure, such as privacy, a concept vital not only in relation to ethics but also one that could be pertinent to the ability of project coordinators to design and execute effective projects.

Recent years have seen a rise in public concern with the privacy and security of data transmitted over the Internet (Ofcom, 2017). In parallel, the new EU-wide privacy laws that became binding in all member states in May 2018, reflect a new regulatory approach to protecting the privacy of individuals, one that emphasises informed consent and clear communication (European Union, 2016). Citizen science has experienced issues in this area. For example, a report investigating how coordinators ask for consent to store, collect and process data – otherwise known as informed consent – showed that 30/118 (25%) surveyed projects did not ask their participants to provide informed consent (Schade & Tsinaraki, 2016). In other words, a quarter of these projects did not take precautions to ensure that any disclosed or collected information provided by citizen scientists was given by individuals who fully understood not only why they were providing information but also how this information would be handled and utilised. Moreover, over half of the surveyed projects (64/118, 53%) only asked for consent with a general Terms of Use (ToU) umbrella-type agreement. Only 10 projects out of 118 asked participants to sign an explicit informed consent form. The remaining 12 projects (14%) used other solutions.

Even when special care is taken to familiarise participants with potential risks to data privacy and measures implemented to counteract them, it is unclear whether participants achieve a working understanding of these issues. Bowser and Wiggins (Bowser-Livermore & Wiggins, 2015) surveyed 30 participatory research projects to see how privacy policies were presented and how they aligned with actual practices. They found that, alongside inconsistencies between policy and implementation, some of the policies were incomplete and lacking in terms of clear communication about what happens to participants' data. As many citizen science projects struggle with recruitment and retention of participants (Riesch & Potter, 2014), it would not be unreasonable to wonder whether a less than rigorous approach to data-use consent (Schade & Tsinaraki, 2016), might be due to a fear of discouraging existing participants from disclosing or supplying data.

However, to ensure that citizen science projects are ethically sound, participants ought to be encouraged to disclose data not via lack of information but rather through honest and transparent communication that realistically describes the benefits of taking part in citizen science. While citizen scientists may decide to share their commodity – personal information – because of the potential gains that participation in a citizen science project might bring them, be it learning, supporting scientific progress or contribution to a common goal (Lee, Crowston, Harandi, Østerlund, & Miller, 2018) – the increased scrutiny, legal and societal, of the ways institutions gather and use sensitive data, requires citizen science coordinators to place more emphasis on informed consent. This, in turn, could lead to greater difficulties with the recruitment of citizen scientists who are willing to share their data.

On the one hand, the seminal work of Westin assumes the existence of three categories of people, in relation to their privacy attitudes. These are the (1) *Privacy Fundamentalists*, (2) *Privacy Unconcerned* and, (3) *Privacy Pragmatists* who, respectively, (1) have a deep concern for privacy, (2) are happy to freely share information, and (3) will share information based on what they are offered in return (Harris and Associates Inc. & Westin, 1998). On the other hand, however, the relationship between privacy attitudes and disclosure behaviour is not straightforward, and fear of a privacy breach does not necessarily result in caution when disclosing information. This phenomenon, referred to as the *privacy paradox*, suggests that other factors are at play when the public makes decisions about whether, or not, to disclose information. Scientists have attempted to explain the privacy paradox via several theories, including those centred on a calculation of risks and benefits (‘privacy calculus’); reliance on psychological heuristics; access to incomplete information; and social motivation (Kokolakis, 2017). Nevertheless, a big chunk of the work examining disclosure of sensitive personal data centres on the creation of theoretical explanations, with experimental studies making up only a small part of research in the area (Gómez-Barroso, 2018).

In a citizen science setting, it is likely that a data disclosure decision could be influenced by several factors simultaneously, and that various theoretical explanations could be applied to shed light on why citizen scientists choose to disclose or withhold data. In practice, however, it is important to study how project designers and coordinators can encourage citizen scientists to share data about themselves, through transparent and ethical means – making projects that rely on the disclosure of personal information sustainable.

As noted by Gómez-Barroso (Gómez-Barroso, 2018), in a data disclosure situation, there are two '*main incentives to which a rational user could respond: a better service, and money.*' (p.1482). Monetary incentives are not, however, considered compatible with citizen science as they conflict with the key principle that citizen science is participation of laypersons in scientific activities for the purpose of learning, engagement and supporting research (Theobald et al., 2015). It is therefore not possible to extrapolate findings from market research directly to the analysis of data disclosure in citizen science. Instead, it is more appropriate to study how non-monetary rewards specific to the context of citizen science may be related to citizen scientists' decisions to disclose or withhold information. However, we cannot expect research into participant motivation to feed directly into the design of citizen science projects that encourage disclosure of data. Motivation in citizen science can vary across levels of commitment (Eveleigh, Jennett, Blandford, Brohan, & Cox, 2014) and it can affect different stages of participation in different ways (Rotman et al., 2012). This is why it is important to experimentally study the relationship between motivation and data disclosure in citizen science.

To summarise, it is imperative to investigate how we can motivate citizen scientists to share their data within citizen science projects through ethical means. Stricter privacy laws necessitate increased clarity in communication with individuals when encouraging them to disclose personal data. This thesis aims to identify which modes of communication about project outcomes will help citizen scientists confidently share personal data and maximise the benefits of participation.

1.2. Research aim and questions, approach and contribution

The overarching aim of this thesis is to answer the question:

How does communication about the rewards of citizen science participation affect the willingness of citizen scientists to share personal data?

To achieve this, studies reported in this thesis focus on investigating the impact of different motivational factors on the volume of personal data disclosed by citizen scientists. In practical terms, we sought to answer the question: how can communication about the benefits of taking part in citizen science encourage participants to confidently disclose personal data? To do this, we employed methodology utilised in research into recruitment for citizen science projects. This included drawing from research that investigates how different messages about the aims of participating in citizen science vary in their effectiveness at recruiting participants.

Specific research questions related to individual studies were as follows:

Study 1

1. Do privacy attitudes account for the pattern of data disclosure among citizen scientists?

2. Can short motivational messages influence the volume of data disclosure among citizen scientists?

Study 2

3. Do citizen scientists’ preferences about the rewards of citizen science influence the volume of data disclosure?
4. Which type of motivational messages about learning opportunities (learning about oneself or learning about science) results in a larger volume of data disclosure?

Study 3

5. Does the way a citizen science project is advertised influence the level of data disclosure?
6. And does this in turn influence the relationship between motivational messages and the disclosure of data?

Study 4

7. Can motivational messages influence the volume of data disclosure in a sample of citizen scientists recruited from an existing online community?

Study 5

8. Can motivational preferences influence the volume of data disclosure in a sample of citizen scientists recruited from an existing online community?

The key contribution of this thesis is an advancement of understanding about the types of benefits that citizen scientists expect to achieve when disclosing personal data. In practice, this translates into a set of guidelines that will help project co-ordinators design effective citizen science programmes. Furthermore, this work helps create an empirical basis to supplement the current theoretical discussions in citizen science on the topics of transparency, and fair exchanges between those who host citizen science projects and those who voluntarily offer their time and labour to such initiatives.

The theoretical contributions include the demonstration, among citizen science dabblers, of: the effectiveness of short motivational messages, presented both pre-recruitment and post-recruitment, in influencing the disclosure of data; the lack of a relationship between privacy attitudes and data disclosure; and the differential impact on data disclosure of prompted and preferred motivation. An important theoretical contribution, with relevance for the wider study of motivation in one-off citizen science projects, is the evidence for the distinction between the motivation to participate and the motivation to disclose data.

This thesis also offers ethical contributions. Evidence for the link between information provided to participants about citizen science projects, and their willingness to disclose data within projects, translates into a responsibility for projects coordinators to communicate with clarity and accuracy. It also further highlights the need for empirical evaluations of citizen science projects, which examine the realistic and concrete rewards of projects for the participants, researchers, and the wider society. Together, clear communication and empirical evaluations, have the potential to shift the focus of citizen science from institutional productivity, to benefitting all stakeholders.

1.3. Thesis structure

This thesis consists of 6 chapters.

Chapter Two draws together the literature on privacy and data disclosure and the literature on citizen science. First, it discusses contexts of data disclosure and outlines the reasons why people share data in daily and citizen science settings. The recent regulatory changes and the increased regulatory emphasis on informed consent are highlighted. We summarise the limited available research on the data disclosure among experienced citizen scientists and discuss how experimental findings from data disclosure research can be applied to the area of citizen science. We also discuss the available research on the rewards of citizen science, and the role of rewards (incentives) in recruiting participants to projects. This is followed by a discussion of the transparency of outcomes in citizen science, and the impact that lack of accurate and realistic communication may have on the data disclosure in citizen science. The final section of the literature review identifies a research gap: casual (inexperienced) citizen scientists stand to be the most influenced by the communication about rewards and incentives of participation, yet no work has been undertaken to understanding the relationship between motivational communication and data disclosure in that population of citizen scientists.

Chapter Three provides a rationale for the method chosen for the studies in this thesis: we explain why we decided to study data disclosure in the format of an online survey. Following a section dedicated to ethical approval, this chapter outlines the research method used in this thesis, describing the recruitment of participants, materials, research design and procedure. The ‘Materials’ section includes a validation study for neutral and sensitive survey questions. In this validation study, we demonstrated that the questions defined as requesting sensitive or neutral data were correctly interpreted by the participants, and therefore appropriate for the use in our studies.

Chapter Four presents three experimental studies conducted in the context of a citizen science survey that explores sleep habits. These three studies demonstrated and explored the impact of motivation to learn on the volume of data disclosed by citizen scientists. In each study, an online citizen science project invited participants to disclose personal data. **Study 1** sought to examine the links between motivation to participate in citizen science and the willingness to disclose personal information in such settings. The results suggest that presenting a motivational message, at the start of a citizen science project that involves the disclosure of personal data, may have an impact on the disclosure behaviour of participants. **Study 2** further explored the effectiveness of motivational messages emphasising Learning, by comparing the disclosure of data across three motivational groups: 1) learning about science, 2) learning about oneself, and 3) a control group presented with a message that welcomed people to the survey. Moreover, by asking participants about which benefits of citizen science should be prioritised in future projects, we investigated the impact of motivational preferences on data disclosure. In this study, neither motivational messages nor motivational preferences could account for the pattern of data disclosure. **Study 3** compared the relationship between motivational messages and disclosure of data across two types of projects: a citizen science survey and an online quiz. We found that the relationship between motivational messages and data disclosure was different across those two types of projects.

Chapter Five presents two studies conducted in the context of a citizen science project aimed at pet owners. These studies achieved two aims. Firstly, they provided an investigation of data disclosure among citizen scientists who (in contrast to participants in Studies 1-3) were recruited from an existing online community. Secondly, these studies examined the effectiveness of two motivational devices, motivational messages, and a motivational question, in the same population of participants. **Study 4** explored the impact of the same motivational messages that were used in Study 1. We found that participants who were presented with the message emphasising Altruism disclosed the largest volume of data, nevertheless, the differences between motivational groups were not significant. **Study 5** employed a motivational question. Our aim was to compare the relative impact of a randomly assigned motivational message (Study 4) and a motivational question (Study 5). We found that the majority of participants in Study 5 chose altruism and contribution as the benefits of citizen science that should be prioritised. However, the less popular benefits of social interaction and learning were associated with a larger volume of data disclosure.

Chapter Six concludes the thesis by directly addressing the main research aim, discussing the findings in the context of the relevant literature and outlining practical guidelines for designers and coordinators of citizen science projects that involve the disclosure of personal data. Special emphasis is placed on discussing the differences between motivating people to participate in citizen science and motivating them to share data within these projects. This chapter also discusses the ethical issues related to studying data disclosure among citizen scientists: recruitment from an existing community, the use of deception and addressing participant complaints, as well as the ethical use of findings reported in this thesis.

Chapter 2

Data disclosure in citizen science projects

2.1. Introduction

This chapter presents and discusses the issues related to data disclosure in citizen science. It begins by describing how the majority of data disclosure now takes place online, both in everyday life and in more specific contexts such as citizen science. This is followed by a discussion of the recent regulatory changes in the area of data protection – we argue that the increased regulatory emphasis on informed consent means that it is important to accurately describe the benefits of taking part in citizen science when asking participants to share personal data.

We outline what makes people decide to share their data online and describe the limited research available on data disclosure in citizen science. We highlight the role of transparency in citizen science and illustrate the potential for miscommunication between project leaders and participants in citizen science. The implications for data disclosure in citizen science are discussed.

The final section of the literature review identifies a research gap: no work has been undertaken to investigate the disclosure behaviour of casual (inexperienced) citizen scientists, who due to their lack of experience stand to be the most influenced by communications from project coordinators about the aims and benefits of taking part in a project.

2.2. Disclosing data online

2.2.1. Disclosure in everyday settings

Recent years have seen the migration of labour, leisure activities, and everyday administrative tasks, both private and professional, to the online Web (Vilhelmson, Thulin, & Elldér, 2017). The degree to which it is now accepted, as well as expected, for individuals to ‘live’ online, has increased immensely. As observed by Hollandsworth et al. (Hollandsworth, Donovan, & Welch, 2017), technologies of the digital world are not only pervasive but also, in a sense, inescapable. It is not possible to ‘go home’ or opt-out of systems embedded into the daily life tasks and communication pathways. While this interconnectivity brings varied benefits, it also carries a multitude of drawbacks such as a precarious relationship between excessive Internet use and mental health as well as risks to psychosocial wellbeing (Caplan, 2003). The array of potential dangers also encompasses a rather more technical subject of data security and privacy. The interconnectivity mentioned above puts at risk the ability of individuals to preserve a sphere of privacy. It becomes more difficult to separate and protect information that we wish to keep confidential for legal, social or other reasons. The same technological advances that enable us to share data online and benefit from it (for example by accessing a bank account without leaving home, or easily communicating with friends via social networks) will also make it more and more difficult to only share those data with our chosen end-point, be it a private recipient, a retailer, or a government body.

Currently, a disproportionate amount of data disclosure takes place online, both passively, via mechanisms such as tracking or browsing logs, but also through active and sometimes extensive disclosure of personal information in everyday situations such as those pertaining to healthcare (Bath, 2008), banking (Chong, Ooi, Lin, & Tan, 2010), insurance (Mackert, Koh, Mabry-Flynn, Champlin, & Beal, 2017), or even psychotherapy (Blackmore, Tantam, & Deurzen, 2008). Similarly, many citizen science projects take place online, and taking part often means that individuals are disclosing some type of personal data, such as their email address, health information, or even location of their home.

2.2.2. Disclosure in citizen science

2.2.2.1. What is citizen science?

Citizen science is the practice of involving laypersons in various stages of research. In contrast to research participants, who typically receive some form of compensation for their time, citizen scientists contribute for the purpose of education, enjoyment, societal contribution and feelings of accomplishment, to list just a few of the typically named benefits of participation in citizen science programmes (Silvertown, 2009). The concept of 'citizen science' is often used as an umbrella term, with the European Commission's Digital Science Unit referring to '*the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources. Participants provide experimental data and facilities for researchers, raise new questions and co-create a new scientific culture*' (Socientize Project, 2013, p.6). In the context of this thesis, we understand citizen science participation as people's voluntary and unpaid participation in scientific activities.

While early citizen science relied on traditional data-gathering instruments, currently participants often use specialised sensors embedded into their phones and citizen science applications that can be accessed through the same device. An increasing number of citizen science projects involve the use of digital devices, which the participants use to collect and upload data, communicate with one another and with project co-ordinators or even find out about participation opportunities, to begin with (Preece, 2016).

2.2.2.2. Citizen scientists and sensitive data

A recent meta-analysis of citizen science projects identified three main categories: the largest one focuses on projects concerned with biology, conservation, and ecology (primarily collection and classification of data); the second category is concerned with citizens collecting geographical data, and the third one with studies in the domains of social sciences and epidemiology (Kullenberg & Kasperowski, 2016). Although the potential for disclosure of personal data exists predominantly in the last area, the first two can often involve the disclosure of time-stamped and location-stamped data, which could pose a privacy risk to participants.

As the array of tools that enable citizen scientists to gather and share data increases, it becomes ever more important to ensure that adequate privacy protections are in place. Nevertheless, it is often a challenge to specify what precisely privacy stands for (Solove, 2006). In an early classification, Goodwin (Goodwin, 2019) highlighted two dimensions of privacy concerns: concern about disclosure of information to those who are not present during the original interaction, and the unwanted presence of others in the individual's physical environment. Data gathered by citizen scientists will often pertain to these individuals' health, homes, family, friends, and sometimes to unrelated passers-by. Such data may reveal consistent patterns of behaviour and situational factors and this, in turn, could pose a risk to citizen scientists' privacy as well as their safety (for example, when an individual posts a location-stamped update, and mentions that they are about to leave home – unaware that this could make them vulnerable to bad actors; (Bowser, Shilton, Preece, & Warrick, 2017)).

Concern for data privacy in citizen science is an issue of particular importance now that digital technologies have made it feasible for projects to recruit immensely large volumes of participants. The eBird programme (Sullivan et al., 2009) is discussed by Bowser and colleagues (Bowser et al., 2017) as an example of a citizen science initiative with privacy risk. The eBird has attracted over 150 000 volunteers from its inception in 2002, and the data disclosed by participants are location- and time-stamped, as well as having the volunteer's name attached to each observation. Such extensive data could easily allow third parties to predict where a particular volunteer, or group of them, might be at a particular point in time. These issues become especially important in light of the recent regulatory changes and the greater public awareness of privacy issues.

2.2.3. Regulatory changes

In recent years, surveys have shown an increase in people's concern about privacy. The 2016 TRUSTe Privacy Index survey has revealed that more Americans worry about online privacy than about loss of personal income, with nearly half of respondents (45%) experiencing more online privacy concerns in 2016 than they did in 2015 (TRUSTe, 2016). According to the Ofcom reports from 2015, 2016 and 2017, a trend for growing privacy concerns among Internet users can also be observed in the United Kingdom (Ofcom, 2015), (Ofcom, 2016), (Ofcom, 2017).

A significant factor influencing both the gathering of data and the attitudes of those who disclose information is the legal and regulatory environment that applies to the collection and retention of data. This typically depends on the geographical location in which the data are gathered (or for online data gathering the location of the user). Regulatory regimes are germane not only in the context of limits placed on gathering, retention, and sharing of data but also in relation to the legally required transparency measures. Although it has been shown that Internet users see the level of information about how their data are shared as insufficient (Balebako, Jung, Lu, Cranor, & Nguyen, 2013), perception of privacy is context-dependent (Poikela & Kaiser, 2017). The existence of legal and other protections may have an impact on the disclosure behaviour of individuals affected by them.

Legally, citizen scientists remain within the same boundaries that have been delineated to individuals outside of any organised programs. Additional requirements may be set by academic ethics committees however citizen science projects are often at least partially co-hosted by non-academic institutions, and the level to which control of data management in real-time can take place will vary across settings. Moreover, the funding bodies subsidising citizen science programmes can also set out additional requirements as to how privacy issues should be communicated to participants and what happens to data after project completion.

It should be noted that recent years have seen an introduction of stricter privacy laws. An example of that is the General Data Protection Regulation (European Union, 2016) binding for all EU member states, as well as the Data Protection Act 2018 (UK Parliament, 2018), introduced in the United Kingdom and modelled after the GDPR. The European Union states that "*The EU General Data Protection Regulation ensures that personal data can only be gathered under strict conditions and for legitimate purposes. Organisations that collect and manage your personal information must also protect it from misuse and respect certain rights*." [1] Efforts to strengthen legal protection for individuals have also taken place in other parts of the world, for example in California (de la Torre, 2018).

Therefore, it is particularly important to examine how citizen scientists disclose or withhold data when faced with new and more detailed consent forms, that are now the legal requirement for projects hosted within the EU and the UK as well as for projects that may recruit residents of those regions. 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 (Marreiros, Tonin, Vlassopoulos, & Schraefel, 2017). While this line of reasoning could motivate project coordinators to avoid

---

[1] https://ec.europa.eu

discussing consequences of data disclosure at length, the new regulatory requirements make it vital to discuss with citizen scientists how their data will be used, stored and processed, and to actively promote fully informed consent. It is, therefore, all the more relevant to study how disclosure of information can be motivated and encouraged even when participants experience greater saliency of potential privacy issues.

2.3. Understanding data disclosure in citizen science: Lessons from privacy research

Researchers have undertaken extensive work examining why individuals decide to share data online despite seemingly high societal levels of privacy concern. Privacy theories, which aim to explain this ‘privacy paradox’, distinguish between a rational weighing of pros and cons of disclosure, and the less conscious scenario of cognitive biases (for example, research conducted by Kehr et al. (Kehr, Kowatsch, Wentzel, & Fleisch, 2015), suggests that when a user interface elicits positive emotion, people underestimate the risk of sharing their data).

The seminal work of Westin suggests that people’s willingness to share information is related to their attitudes to privacy, proposing 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), (Harris and Associates Inc. & Westin, 1998). It has been argued that attitudes to privacy cannot reliably predict privacy behaviours (Woodruff, Pihur, Schmidt, Brandimarte, & Acquisti, 2014). Attitudes may have a mediating role: in a recent study, where participants were told that they were filling a credit card application, individuals who were *not* classed as *Privacy Fundamentalists* (i.e. either *Privacy Pragmatists* or *Unconcerned*) according to the Westin Privacy Scale, were 5.6 times more likely to submit the form (Malheiros, Brostoff, Jennett, & Sasse, 2013).

2.3.1. Experimental study of disclosure behaviour

The study of when and why people are willing to disclose personal data often relies on an examination of their intention or attitudes and rarely focuses on actual behaviour (Barth & de Jong, 2017). As noted by Gómez-Barroso (Gómez-Barroso, 2018), despite the abundance of theories attempting to explain human behaviour in data disclosure situations, there are disproportionately few studies exploring disclosure experimentally. Gómez-Barroso classifies the existing experimental research alongside five categories: experiments on the existence of the privacy paradox (the discrepancy between attitudes and behaviour); experiments on the importance of trust and control as mediators of self-disclosure; experiments on the impact of incentives (either monetary or in the form of customized products); experiments on the willingness to pay for privacy protection; and experiments on nonrational behaviour.

Of those, most relevant to current research is the study of the impact of incentives. Interviews with long-time citizen science participants suggest that it is the benefits of participation that make them choose to contribute data, even when they have concerns about privacy. Moreover, the ethos of citizen science requires a two-way exchange between project coordinators and participants. Investigating whether the promises made by project coordinators about the benefits of taking part in citizen science can have an impact on how much data people disclose will help make such exchanges more transparent and fairer.

2.3.2. Motivation to disclose data

In experimental settings, researchers typically study people’s willingness to part with their data due to two types of incentives: monetary or personalisation (Gómez-Barroso, 2018). The prospect of a customised service can sway participants towards disclosure, but it can also make them wary and privacy-conscious (Ward, Bridges, & Chitty, 2005). Moreover, the impact of personalisation becomes weaker when the sensitivity of data increases (Mothersbaugh, Foxx, Beatty, & Wang, 2012). Many studies include cash incentives or vouchers as incentives to share data. While overall seen as effective at eliciting disclosure, cash presents its problems, with some individuals using deception to receive a monetary incentive without sharing their real data (Steinfeld, 2015).

It is not possible to directly extrapolate from the wider arena of privacy research, typically conducted in the context of consumer behaviour, to the area of citizen science. The ethos of a voluntary contribution to citizen science projects makes it unsuitable to offer cash incentives. However, there is evidence to suggest that the same types of motivation that encourage people to take part in citizen science may also prompt them to share information despite having privacy concerns.

In the first empirical investigation of data disclosure in citizen science settings, Bowser and colleagues (Bowser et al., 2017) conducted interviews with 14 experienced citizen scientists and 13 citizen science coordinators. Interviewees spoke at length about the motivational factors associated with participation in citizen science, such as being able to contribute and see other people's contributions (volunteers in this study enjoyed comparing their contributions against other people's data), eagerness to learn, and social opportunities (one of the citizen scientists suggested that collecting data for citizen science alongside friends can feel like a party). The study found that the motivational factors that attracted and retained participants in citizen science projects were strong enough to overpower any of the privacy concerns that these individuals might have had about sharing their data and metadata. Bowser et al.'s participants demonstrated only limited concerns about privacy, perceiving scientists who coordinated projects as *'trusted* experts' (p.2131). The concerns they did have exerted a negligible impact on their willingness to disclose information, suggesting that privacy concerns in the context of citizen science are not a risk to participation when the motivation to take part is strong.

The authors discussed their findings in relation to Nissenbaum's contextual integrity framework. According to Nissenbaum (Nissenbaum, 2010), the context of data disclosure has a major impact on how an individual understands and perceives data privacy. This context is made up of situational factors such as the type and value of data, and how and why data are shared, as well as for what purpose. These factors then have an impact on whether an individual decides to disclose or to withhold information. Bowser et al. found that the level of concern with privacy was relatively low and even when present, such concern did little to stop participants from disclosing data requested from them: in large part, the specificity of the citizen science context, which underscores principles of sharing, co-creation and institutional trust, encouraged disclosure.

An important link was made between motivation for taking part in citizen science and a willingness to share personal data. The key motivations for participants of this study were the desire to share, personal interest (such as learning), and the broader concept of curiosity. As emphasised by the researchers, data sharing was also likely to have been encouraged by some of the place-based motivational aspects, such as wanting to be part of the local community or to collaborate on a project with like-minded individuals (Bowser et al., 2017).

Most studies that look at how rewards impact the disclosure of personal data do so in the context of the private sector. Private companies can incentivise people to share their data by offering a better service or a better deal – In experimental settings, researchers typically study people's willingness to part with their data due to two types of incentives: monetary or personalisation (Gómez-Barroso, 2018). We cannot realistically offer monetary or personalisation incentives in the context of citizen science, due to its voluntary and non-commercial nature. What we can do, however, is highlight the rewards intrinsic to citizen science participation, so that they have an active role in the decisions that citizens make about the disclosure of information.

2.3.3. What are the rewards of taking part in citizen science?

In contrast to traditional research, or contexts in which people's data are collected by private companies, citizen scientists offer their time and effort, and, frequently, information about themselves, on a voluntarily basis. Citizen scientists do not receive any monetary reimbursement nor other material incentives such as discounts on services or entry into prize draws. Instead, they benefit from non-material rewards, for example learning about their area of interest, acquisition of skills, feeling of contribution, discovery, or enjoyment of the task (Raddick et al., 2013). In the context of this thesis, the term 'reward' refers to what the citizen scientist receives as a result of taking part in a project. In other words, the reward of taking part in citizen science projects is what motivates a particular individual to take part, despite the lack of monetary or other material incentive. Project coordinators often communicate about the rewards of participation at the point of recruitment, in order to attract and retain participants.

2.3.3.1. Communicating rewards to aid participant recruitment

Motivation to participate is one of the most widely researched aspects of citizen science (Diner, Nakayama, Nov, & Porfiri, 2018). When discussing motivation to take part in citizen science, researchers make a distinction between *intrinsic motivation* related to the act of participating (such as fun, avoidance of boredom, enjoyment of social interaction) and *extrinsic motivation* related to the expected project outcomes (for example, wanting to support scientific discovery or acquire skills). Studying participation patterns, Eveleigh et al. (Eveleigh et al., 2014) observed that citizen scientists who care more about intrinsic benefits of participation are more likely to stay loyal to a project and contribute in greater volume, while the more casual contributors care about extrinsic benefits and wish to stay informed about the outcomes of projects.

The unifying motive of studying participant motivation is the desire to channel this knowledge into better recruitment and retention strategies. Studies have shown that the information presented to potential participants at the point of recruitment can have a significant impact on whether they decide to join a project, and how much time they are then willing to contribute. A recent experiment that looked at the effectiveness of email messages which 'frame' a citizen science project in a particular way demonstrated that reframing recruitment slogans from 'contributing to science' to 'comparing oneself to others' can increase recruitment substantially, in this case by nearly 9% (August, Oliviera, Tan, Smith, & Reinecke, 2018). Importantly, this study was conducted among LabintheWild participants (https://labinthewild.org is a platform that hosts psychological citizen science experiments), that is individuals who took part in citizen science in the area of psychology, contributing data about themselves.

A recently conducted large study (n = 36,513) looked at the effectiveness of attracting individuals to a citizen science project by varying the subject line of an email and emphasising one of the four motivations: learning, social proof, contributing, or altruism (Lee et al., 2018). Potential participants were sent emails with one of the four randomly assigned messages in the subject line. The four variations of the message all aimed 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 of gravitational waves!*); 4. Altruism ('*Astrophysicists need your help to classify problematic noise in the search for gravitational waves!*'). Lee et al. found that different motivations supported different aspects of participation: while emphasising altruism at the recruitment stage helped attract participants who contributed the greatest volume of work, the most effective at eliciting initial participation from the biggest sample of individuals were messages that focused on contributing to science and learning about science. This study, however, was conducted in the context of the Zooniverse platform (https://www.zooniverse.org), which hosts projects that typically do not require the sharing of extensive personal data, and therefore it is uncertain to what extent results from this study can be extrapolated from to citizen psych-science.

The studies described above demonstrate that different types of citizen science incentives advertised in recruitment messages can have a differential impact on people's willingness to join a project. Therefore, we consider such messages to be an appropriate tool to highlight the benefits of citizen science participation in the effort to encourage the disclosure of personal data by citizen scientists. For studies reported in this thesis, we adapted the recruitment messages used by Lee et al. (Lee et al., 2018). This allowed us to design a set of four motivational messages as well as a motivational question with four answer options – two motivational devices that are further described in the second part of this chapter.

In Studies 2 and 3 we further explored the impact of two facets of the motivation to learn: learning about science and learning about oneself. Section 2.3.3.2. provides background on the importance of these motivations for citizen science.

2.3.3.2. Learning about science or self-discovery?

2.3.3.2.1. Two types of learning motivation

The ability to learn about the scientific process or an area of interest is considered an important benefit of participation in citizen science ((Theobald et al., 2015); (Domroese & Johnson, 2017); (Oberhauser & Lebuhn, 2012); (Bonney et al., 2009)). On the other hand, a large study of the effectiveness of recruitment messages has highlighted the role of the 'learn about yourself' framing, which can help increase the number of people who are willing to take part in citizen science (August et al., 2018). The latter type of learning is similar to the 'quantified self' movement, where laypersons gather data about themselves in the quest for self-knowledge. Wiseman et al. (Wiseman, Cox, Gould, & Brumby, 2017) define citizen science as distinct from the 'quantified self' movement, where '*Unlike CitSci, the motivation for collection of data comes not from the "selfless" act of participating in an experiment belonging to someone else, but rather from the opportunity to learn more about oneself*' (p.2).

One of the difficulties of comparing citizen science projects is that, despite several classifications and taxonomies, there are no specific guidelines on what citizen science can or cannot be. Recently, many such initiatives, especially those in the area of psychology, have used the incentive of finding out something about yourself, to recruit participants (e.g. LabintheWild platform). While participants are not necessarily required to self-track (gather data about themselves), they are encouraged to take part on the basis that they will expand their knowledge about themselves. This could make those studies appear more akin to popular online quizzes than research projects.

The participants' motivation is a central aspect of citizen science because traditionally, citizen scientists do not get paid, while lab-based study participants do typically receive reimbursement for their time. Should we then also distinguish, further focusing on participant motivation, between different types of unpaid research projects? Can online surveys that focus on recruiting people by offering them an opportunity to learn something about themselves still be considered citizen science? These issues should be studied experimentally in order to provide insights about the impact of these two facets of learning motivation on the disclosure behaviour of citizen scientists.

2.3.3.2.2. Is there a need to separate the concepts of an online quiz and citizen science?

Some citizen science projects present themselves in a way that is akin to online quizzes, that is by emphasising opportunities to test yourself or to learn something about yourself (often by discovering a personal score). However, online quizzes have recently attracted some negative attention. For example, a company called company Cambridge Analytica is believed to have gathered data through an online personality quiz ‘thisismydigitallife’, which then allowed it to deliver targeted political advertisements to Facebook users as well as their friends who never took the quiz themselves (Cadwalladr, 2018). This excessive access to personal data was later described as a ‘breach of trust’ by the CEO of Facebook, Mark Zuckerberg (Zuckerberg, 2018). As noted by Paul Bisceglio in an article published in The Atlantic ‘*online personality analysis can easily blur the line between opting in and opting out*’ (Bisceglio, 2017). Another example of the type of online quizzes that extract a lot of information in a way that may not be clear to all users is the immensely popular quiz format used by the Buzzfeed website (https://www.buzzfeed.com). Some of these Buzzfeed quizzes appear more innocuous (for example: ‘Customize Your Cheesecake And We'll Reveal Which High School Clique You Belong In’ (Aspler, 2019)) while others include very invasive questions (the ‘How Privileged Are You?’ quiz asks people to check off all the statements that apply to them: including ones such as ‘I have never been raped’ or ‘I have never attempted suicide’ (Thomas & Jha, 2017)). Such data, sometimes supplemented by other information like the person's gender or email address, are collected by the website for research and marketing purposes; while the data are said to be anonymised and aggregated, some commentators still consider them a cybersecurity risk (e.g. https://barker.co.uk/buzzfeediswatching). Private statements of many websites that provide such 'learn something about yourself' quizzes reserve the right to aggregate data and share them with third parties (https://www.mic.com/articles/188678/its-not-just-facebook-online-quizzes-also-collect-data-about-you). The market-driven approach that prompts private companies to host quizzes online in the effort to extract as much data as possible from the people who fill them, is very different from the atmosphere of trust and co-operation between citizen scientists and project hosts (Bowser et al., 2017) that enables the sharing of sensitive data in citizen science.

2.3.4. The role of transparency in data collection

A major factor influencing the willingness of an individual to share their personal data online will be the degree to which they feel they can trust the institutions/s requesting and handling information. Lack of trust may cause one to feel protective of their data; the development of trust will, in turn, decrease privacy concerns (Taddei & Contena, 2013). Not all institutions or individuals who gather people's data, however, exercise the measures necessary to earn the trust of their data subjects (a term understood here as those about whom data are gathered, or whose metadata are retained).

In privacy research, transparency refers to the degree to which individuals can access information collected about them, as well as the openness on the part of data collectors about why and how these data will be used. Lack of transparency features can lead individuals to withhold their data, as a means of trying to control their privacy boundaries. On the other hand, higher transparency can increase the fear of disclosure by increasing the saliency of privacy concerns (Karwatzki, Dytynko, Trenz, & Veit, 2017).

There are several reasons why individuals may be asked to enter their information online, and why obtaining data may be prioritised over clear communication about the aims of data collection. Data broker companies will often treat privacy issues as secondary since anything that can potentially increase the caution of data subjects will also have the ability to lower these companies' income. As a lot of data collection happens without the knowledge or explicit consent of the individuals about whom data are collected for a particular data broker to use and gain profit from such information – data brokers will often lack an incentive to maintain reciprocity with their data subjects (Crain, 2018). Crain argues that individuals and companies, which stand to gain from ownership of and/or access to personal data, will often avoid alerting data subjects to privacy issues. He posits that '*commodification of personal information lies at the root of the power imbalances that transparency-based strategies of consumer empowerment seek to rectify*' (p.2).

Importantly, limited transparency has the potential to undermine trust: if the institution gathering the data is not clear about what they are going to do with the data then are they trying to hide something?

2.4. Outcome transparency

2.4.1. Why outcome transparency matters for citizen science projects

As highlighted in section 2.3.2., in privacy research, transparency refers to the degree to which individuals can access information collected about them, as well as how these data will be used. As shown by a recent report looking at citizen science projects, project coordinators do not always assume the need to seek informed consent from participants (Schade & Tsinaraki, 2016), possibly perceiving citizen science as having less need for careful consideration of ethics than traditional research.

There is another aspect to transparency, particularly complex in the context of citizen science: being transparent about what an individual can gain by taking part in an activity that requires the sharing of one's data. Incentives offered in the area of citizen science are hard to verify. For example, when someone willing to learn about science is told that a citizen science project will help them acquire new skills, it is hard to assess whether and to what extent this has taken place. In contrast, a private company offering a 10% discount for signing up for a loyalty program has to deliver exactly what they promised.

The General Data Protection Regulation (European Union, 2016) and the Data Protection Act 2018 (UK Parliament, 2018) put an increased emphasis on consent. The link between motivation and willingness on the part of citizen scientists to disclose data, described in section 2.3.1., suggests that presenting inaccurate or unrealistic promises when recruiting people for a citizen science project may not only be unethical but also potentially in breach of the new, stricter privacy laws. However, the area of citizen science is unregulated and some of the claims made by project coordinators may be subjective as opposed to being grounded in empirical evidence. Below, we describe an example of difficulties with maintaining outcome transparency in online citizen science projects.

2.4.2. Are we increasing access? An example of difficulties with outcome transparency

In recent years, multiple institutions have embraced the concept of delegating small tasks to online crowds for cost-effectiveness, speed, and volume of accomplished work, as well as the associated engagement of the public. Online 'crowds' will typically work on small tasks, for example, producing metadata such as tags, ratings or descriptions. This type of contribution fits the definition of volunteer thinking, a type of citizen science task, as classified by Haklay (Haklay, 2015).

Benefits from developing citizen crowdsourcing projects for their hosts, that is universities, libraries, archives, museums, and other predominantly public sector institutions, appear to be self-evident. Hailed as '*productive engagement*' (Ridge, 2013, p.442), citizen crowdsourcing can help accomplish large volumes of output. Productivity is frequently mentioned in the associated literature: top contributors are praised for it (Leon, 2014) and it is used to justify the effort of maintaining crowdsourcing communities (Phillips, 2014).

However, the concrete benefits of such initiatives for participants are less clear. Importantly, in crowdsourcing projects, citizen scientists are often tasked with small and repetitive tasks. They help ease the institutional backlog of work, but they do not have much input into the aims and impact of projects. This is in stark contrast to community-focused citizen science which takes on real-life issues of communities and helps address them by teaching citizen scientists the necessary skills (an example is the Extreme Citizen Science practice (Stevens et al., 2014) or projects tackling environmental issues (Dutta et al., 2009).

Why is it important to keep track of the promises made by project coordinators to the individuals who take part in citizen crowdsourcing initiatives? Citizen crowdsourcing is a type of citizen science participation that will often involve a large number of participants, many of whom are unlikely to have spent much time researching the background of the projects they are taking part in. These 'dabblers' (Eveleigh et al., 2014) either prefer to participate casually or may dabble in several projects before becoming loyal to one of them. Research has found that despite their casual mode of participation, 'dabblers' are particularly interested in the outcomes of the projects they get involved in (Eveleigh et al., 2014). It follows that when recruiting dabblers, they could be especially influenced by the promises that are made to them about what these projects can accomplish.

One of the often-mentioned benefits of citizen crowdsourcing projects is increased access to research in areas of science and humanities, as well as access to heritage collections such as archives and digitised museum artefacts. But do project coordinators emphasise increased access for diverse populations or do they prioritise institutional benefits such as easing the backlog of work? Emphasis on maximising output from citizen crowdsourcing leads researchers to recommend niche-sourcing, which means targeting individuals with skills and interests most suitable to one's projects (Oosterman et al., 2014). If that is the case, can we truly speak of an increase in access? Or are we simply looking to involve a large number of individuals with existing expertise and skills in unpaid labour?

Another argument used in the literature to demonstrate improved access is that involving the volunteer 'crowd' in the tasks of heritage institutions such as museums and archives can help democratise cataloguing systems. Delegating the creation of structured metadata used in cataloguing digital items, to online volunteers has been described as a '*cost-effective means of increasing access*' (Dunn & Hedges, 2012, p.28). Metadata authored by the public (rather than by professional staff) are thought to result in democratised cataloguing systems. For example, crowdsourced tags are meant to better represent vocabulary of the wider society, thus enabling easier searches (Eccles & Greg, 2014): *"taggers' terms are perhaps more likely to be used by other people of similar backgrounds when looking for pictures with those attributes. Thus social tagging provides a set of search terms designed by, and for, the widest set of potential users*' (p.190). However, it is unclear whether involving citizen scientists in creating tags does result in improved searches. In a recent study of language behaviour, tags provided by online crowds included a considerable degree of potentially unhelpful gender-related bias, for example when tags were created for men and women depicted within occupational roles, pictures of women received more tags relating to appearance, while occupation-related tags were more frequently given to men (Otterbacher, 2015).

Another barrier to access is the overrepresentation of skilled super-contributors in some projects. It is noteworthy that only 3% of *Transcribe Bentham* participants were not educated to at least undergraduate level, and that '*almost a quarter achieved a doctorate*' (Causer & Wallace, 2012, p.13). Ability and willingness of diverse populations to access digital resources will be limited by contextual factors, such as lack of equipment or the digital divide – a well-studied phenomenon that pertains to how a person's behaviour, skills, habits, beliefs about one's ability or self-assessed worthiness will be as impactful online as they are offline (Pearce & Rice, 2013). Are audiences more diverse, or do the same/similar audiences gain access to larger databases? The practice of targeting well-educated individuals to improve the quality of crowdsourced data (Oosterman et al., 2014) suggests that there is a risk of overlooking, or perhaps neglecting underprivileged populations.

With the growing societal and regulatory emphasis on informed consent, it is becoming increasingly important for citizen science coordinators to only make the promises they can fulfil, particularly in cases when the projects involve the collection of personal data. It is also important for researchers to study the relationship between the incentives offered at the point of recruitment to citizen science projects and the willingness of citizen scientists to disclose personal data. If the links between promised benefits and data disclosure can be experimentally evidenced, this will have implications for the design of projects that are both effective and compliant with data protection laws.

2.5. The knowledge gap: What motivates ‘dabblers’ to disclose data?

The issues with outcome transparency described above are particularly pertinent to casual or inexperienced citizen scientists. A lack of extensive knowledge about citizen science could make those individuals rely on the information they receive at the point of recruitment. It makes them more vulnerable to inaccurate or overblown promises that could influence how they make their decisions about whether to disclose or withhold data within projects.

Notably, we do not know much about how casual citizen scientists make decisions about data disclosure. The only empirical study enquiring into information disclosure in citizen science published prior to our research focuses on a sample of committed contributors, that is, participants who demonstrate a high level of knowledge and engagement in citizen science (Bowser et al., 2017). As noted by the authors, recruiting through an event at a natural history museum allowed them to recruit ‘*particularly engaged volunteers familiar with the culture and norms of citizen science*’ (p. 2127). Those individuals who participate more casually (or even for the first time) and therefore have lesser understanding, knowledge, and arguably lower commitment to specific projects, were not included in Bowser et al.’s (Bowser et al., 2017) sample. The finding that citizen scientists disclose information despite privacy concerns is not, therefore, necessarily applicable to all citizen scientists.

Similarly, most studies that explore what motivates citizen scientists to take part in projects rely on self-selecting samples of the most dedicated contributors, who are already so engaged with the projects that they are also willing to take part in surveys about these projects. This is problematic, as it is not the committed long-term contributors but the ‘long tail’ of casual ‘dabblers’ (Eveleigh et al., 2014) that provide the largest volume of volunteer effort (Ponciano, Brasileiro, Simpson, & Smith, 2014). In effect, we neither understand how and why casual citizen scientists (dabblers) disclose data nor is there much research into the motivation of dabblers that we could extrapolate from.

There is evidence to suggest that participant motivation will vary between casual and committed (experienced) citizen scientists. Eveleigh and colleagues (Eveleigh et al., 2014) demonstrated that within the *Old Weather* citizen science project, both intrinsic (related to the enjoyment of activity) and extrinsic (related to the expected outcomes of activity) motivation were positively correlated with the number of contributions, but only the extrinsic motivation was positively correlated with depth of participation (e.g. writing forum posts or transcribing event information). These results highlight that, despite only occasional participation, the casual ‘dabblers’ do care about the outcomes of projects they contribute to.

Casual contributors will often have limited knowledge about a project before deciding to take part. Likely, the features of project design (for example whether the benefits of taking part are advertised or not) will have a meaningful impact on whether these individuals share or withhold data. Establishing a connection between the motivation of a citizen scientist to participate in a project and their willingness to disclose personal data is an important step towards discovering how to encourage sufficient numbers of participants to enter citizen science projects and then to disclose sufficient volume of personal data to make these projects successful. In the case of dabblers, the lack of long-term attachment to a project makes it impossible to study their pre-existing motivations at length. However, experimental manipulations that influence citizen scientists’ motivation, directly before they take part in a project, can be used to study the relationship between motivation and data disclosure in the dabbler population.

Moreover, some initiatives uniquely depend on one-off participation of casual contributors; this is especially the case with psychological and health surveys, where the citizen scientist is simultaneously the study subject (Jennett et al., 2014). It is important to investigate how features of one-off citizen science projects might impact the degree to which participants disclose or withhold information.

2.6. Summary

To summarise, in the participatory framework of citizen science, disclosure of information can happen as a result of a decision on the part of a participant to offer their data in exchange for the benefits that programme participation brings them. It follows, therefore, that what citizen scientists are told about the rewards of taking part in citizen science may sway them to disclose or withhold data. This becomes problematic when project leaders offer unclear or overly optimistic promises about what citizen science projects can achieve. Although citizen science projects involve the collection of various types of personal data, the discussion of privacy issues has so far been mainly theoretical, and this area lacks experimental research. The new privacy laws will require citizen science coordinators to put more emphasis on informed consent and transparent collection of data. Accurately communicating about the concrete rewards of taking part in citizen science will become increasingly important, as project coordinators face the challenges of gathering sustainable levels of data at a time of increased legal and public scrutiny over the ethics of data collection and processing.

Chapter 3

Research method used in this thesis

3.1. Introduction

Chapter 2 outlined how the benefits of sharing information can influence whether an individual decides to disclose their data. For example, the promise of a more personalised service or a cash discount may convince a consumer to share their data with a company. As demonstrated in Chapter 2, individuals who participate in citizen science are not always presented with clear and transparent information about the advantages of taking part in citizen science projects. It is important to examine whether the promises that citizen science coordinators make to their participants – about the expected rewards of participation – can influence the data disclosure decisions made by these citizen scientists. This thesis seeks to identify how communication about the benefits of citizen science can influence the volume of data that citizen scientists disclose.

To achieve the above aim, we conducted five studies, all of which followed the same paradigm: participants were asked to complete an online citizen science survey that included a motivational device and two sets of questions (neutral and sensitive). First, a motivational device either prompted participants to focus on a specific reward of participation via a motivational message or identified participants' preferences via a motivational question. Then, participants were asked to answer a set of Neutral Items (questions that requested data defined as non-sensitive), followed by a set of Sensitive Items (questions that requested data defined as sensitive).

This chapter consists of two parts. **In the first part of this chapter**, we explain why we decided to study data disclosure in the format of an online survey – as opposed to other types of citizen science contributions such as submitting a photo or an ecological observation or completing a classification task.

In the second part of this chapter, we outline the research method used in this thesis, describing the recruitment of participants, materials, research design and procedure. The ‘Materials’ section includes a validation study for neutral and sensitive survey questions. In this validation study, we demonstrated that the questions defined as requesting sensitive or neutral data were correctly interpreted by the participants, and therefore appropriate for the use in our studies.

3.2. The rationale behind the method used in this thesis: The choice of citizen science project type (format of contribution)

In this section, we explain why an online survey was the type of contribution format that we chose for studying data disclosure among citizen scientists.

Studies reported in this thesis examine the disclosure behaviour of citizen scientists in the context of one-off citizen science surveys. While there are different types of citizen science projects that involve the sharing of personal data (for example, conservation citizen science projects ask participants to provide ecological observations accompanied by time, date, and location), the specific type of contribution we chose to study was an online survey that asks participants to provide data about themselves. Firstly, overt disclosure of information about oneself is a feature of the increasingly popular citizen psych-science projects (citizen science projects in the area of psychology). Examples of such projects are Mappiness (https://www.mappinessapp.com), which asks volunteers to record their mood in different locations to discover how different environments affect people’s happiness, and Errordiary (www.errordiary.org), where participants report their everyday errors on the social platform Twitter to help scientists understand the types of errors that people deal with in their lives (Jennett et al., 2014). A hugely successful *Patients Like Me* project, advertised on the SciStarter citizen science platform (https://scistarter.org), encourages people to track their symptoms and connect with others suffering from the same condition. Psychological and health-related citizen science surveys are likely to inquire into topics such as physical and mental health history, and other types of information that are legally classified as sensitive personal data. For that reason, we consider it particularly important to study data disclosure behaviour in this context.

Secondly, the use of an online survey ensures that every participant faces the same disclosure decisions – a standardisation that would be hard to maintain if we studied a different contribution format, for example, the provision of photographs or observations – and therefore allows for a clear-cut comparison of disclosure behaviour across participants.

It should be noted that, in the area of privacy research, surveys are often used to study people’s attitudes, preferences or hypothetical willingness to disclose data. They are sometimes criticised for not investigating authentic behaviour and not allowing for inference of causation. In the context of studies presented in this thesis, however, the online survey was employed to mimic citizen science settings in which participants join a project and share information about themselves. Therefore, in this context, the survey studies actual behaviour and not preferences or hypotheticals.

3.3. Ethical approval

The research presented in this thesis received ethical clearance from the departmental ethics committee (UCL Research Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka). The ethics committee was made aware of the fact that the studies reported in this thesis would use deception: specifically that participants would be told that they are participating in a citizen science project (for example, one investigating the links between stress and sleep) and that it would not be specified that the study is in fact an experiment examining how individuals disclose information in citizen science settings.

Measures were undertaken to ensure that the use of deception was ethical. Firstly, in each study, the Participant Information Sheet (see Appendix A) included a section titled 'Deception', which informed participants that:

'Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).'

Following completion of the study, participants were presented with a debriefing message which explained the aims and methodology of each study. Again, they were reminded that they could withdraw their data and/or contact the researcher with any questions that they might have. Our approach to debriefing evolved over time. At first, we only presented the debriefing message at the end of the study. Following completion of Study 5 (chronologically the second survey to be run), we decided to also ensure that, additionally, each participant who disclosed their email address, would receive the debriefing message via email. Debriefing messages were emailed to Study 5 participants approximately two weeks after the end of data collection. In contrast, participants in Studies 2, 3, and 4, received the debriefing message via email within four hours of taking part in the survey.

Both before and after completion of the survey, all participants were made aware that they could withdraw their data from the study without having to provide a reason. Where requests were made to withdraw data, they were dealt with promptly. It was typical for five or fewer participants per study to contact us and ask for their data to be withdrawn. Only in the case of Study 5, initial five requests to withdraw made before the end of the study were followed by an additional seven requests. The additional requests took place two weeks following the end of the study, shortly after the participants received an email with a debriefing message (only in the case of Study 5, participants received a delayed debriefing message email and more information on this is given in section 3.4.1.2.). It is possible that participants pay more attention to a debriefing message when it arrives with a delay – and more critically evaluate their willingness to take part in a study from the perspective of time. This should be examined in future research as it may have implications for consent procedures in scientific research.

Ethical issues associated with the use of deception in data disclosure research in the context of citizen science are also discussed in Chapter 6.

3.4. Method

Section 3.2. outlined the reasons behind choosing to study citizen scientists' disclosure behaviour in the format of an online survey. This section provides an overview and justification for the broad approach taken across the studies. Details and any variations will be dealt with in the appropriate sections of Chapters 4 and 5.

3.4.1. Recruitment of participants

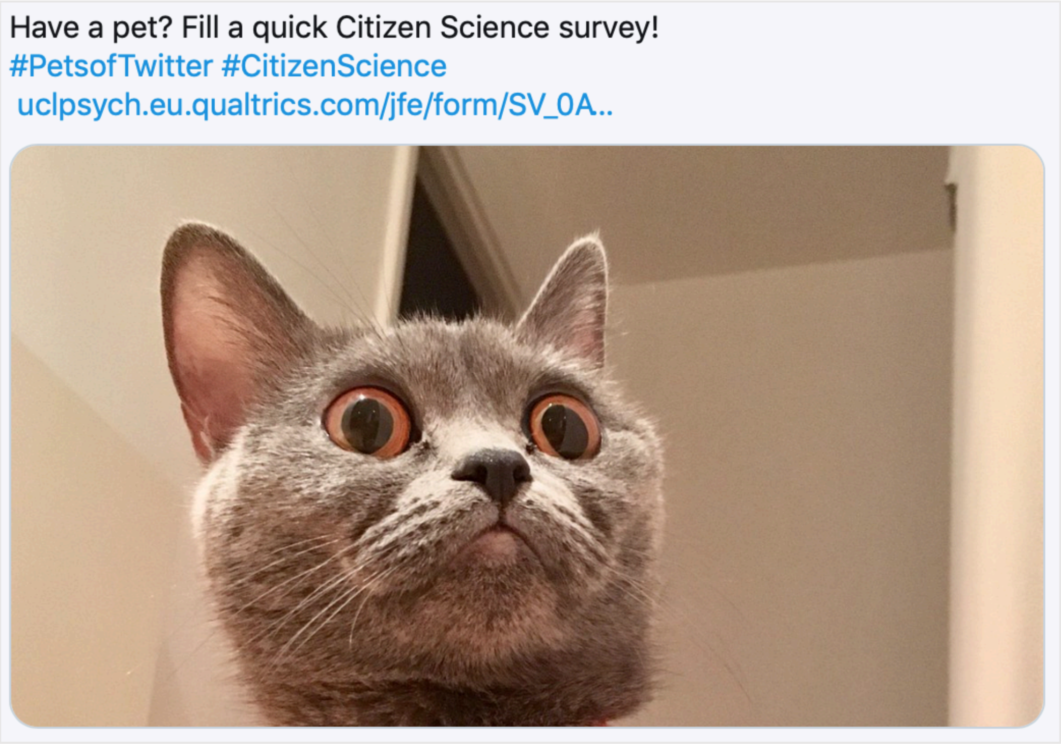
3.4.1.1. General information about participant recruitment

To emulate the authentic citizen science context, we created five fictional citizen science projects for the five studies reported in this thesis. In Studies 1, 2, and 3, participants filled an online survey on the theme of sleep habits. In Studies 4 and 5, participants filled an online survey on the theme of pet ownership. Within these five studies, the recruitment of participants took place online, via Twitter, with participation open to all individuals over the age of 18, irrespective of their location of residence. We created two dedicated Twitter accounts that corresponded to the themes of the survey (sleep habits or pet ownership). These two themes were chosen due to personal interest of the research student – existence of some previous interest in the area was important as it would help sustain the online presence on Twitter and generate Tweets, as well as gather followers, over an extended period of time. Tweets that advertised the link to the survey were amplified through the use of paid Twitter advertisements. These advertisements targeted individuals over the age of 18 who had used the hashtag #CitizenScience as well as theme-relevant hashtags such as #sleep (Studies 1, 2, and 3) or #PetsOfTwitter (Studies 4 and 5).

When composing recruitment Tweets, we ensured the exclusion of any mentions of the motivations that were studied experimentally (learning, altruism, contribution, and social proof). We encouraged people to '*take part in a citizen science survey*' and drew people's attention to those tweets by illustrating them with photographs of animals (e.g., see Figures 1 and 2).

Are you dreaming of a nap?💤😴 Take part in our citizen science 🧪🔎👩‍🔬👨‍🔬 sleep survey! uclpsych.eu.qualtrics.com/jfe/form/SV_8H..

Figure 1 Example of a recruitment tweet (sleep habits context)



Have a pet? Fill a quick Citizen Science survey!
#PetsofTwitter #CitizenScience
uclpsych.eu.qualtrics.com/jfe/form/SV_0A..

Figure 2 Example of a recruitment tweet (pet owner context)

3.4.1.2. Chronological order of participant recruitment

In regard to the organisation of studies within this thesis, studies are grouped according to the two themes: firstly, the sleep habits studies, and then the pet owner studies. Studies 5 and 4 are presented in the reverse chronological order so that both the sleep theme and the pet owner theme chapter would begin with an exploration of the impact of motivational messages.

Below, we provide details about the chronological order in which Studies 1-5 were conducted. Alongside this, we explain how the order in which the studies were conducted has impacted the choice of Independent Variables.

In the first instance we conducted Study 1, which explored the variables of Motivation and Attitude. The variable of Attitude was not found to be responsible for the pattern of data disclosure, and also participant attitudes to privacy were not the primary focus of this thesis. As a result, we decided to discard the variable of Attitude following Study 1. Secondly, we conducted Study 5 in order to investigate data disclosure among citizen scientists in a community setting – we employed the independent variable of Preference, to examine the relationship between citizen scientists’ motivational preferences and their data disclosure behaviour. Following these two studies, we decided to further explore the impact of motivation to learn (a direct continuation of Study 1) and to study the impact of participants’ motivational preferences, this time in the context of a non-community sample – this was done in Study 2. Both Study 5 and Study 2 showed a lower overall mean disclosure than Study 1. These lower mean disclosure scores appeared to be linked to asking people about their motivational preferences (and therefore making them evaluate their willingness to participate more critically). Nevertheless, we needed to rule out the possibility that the participants in Study 2 (where the most popular recruitment tweet did not mention citizen science) may have treated the project as an online quiz and not a citizen science initiative. This possibility was indeed ruled out after Study 3 compared the patterns of data disclosure between people who join an online quiz and people who join a citizen science project. Finally, to make conclusive statements about the respective impact of motivational messages (Motivation variable) and motivational preferences (Preference variable), we needed to have explored both of these variables in the same population of participants. This is why we designed Study 4 to examine the role of the same four motivational messages as employed in Study 1, in the context in which we conducted Study 5.

More detailed discussion on why specific Independent Variables were explored further following Study 1 is included in the Introduction section to each of the further studies.

It should be noted that, as is traditionally the case with citizen science samples, our recruitment relied on volunteer participation. In each of the studies, we continued recruitment until no new participants were joining the project (with the exception of Study 3, where care was taken to recruit matching numbers of participants for each of the two conditions).

3.4.2. Materials: The online survey

3.4.2.1. Survey structure

Each survey in Studies 1-5 was hosted on the Qualtrics platform (https://www.qualtrics.com). Figure 3 depicts, collectively, all elements of the online survey that were used in the five studies reported in this thesis.

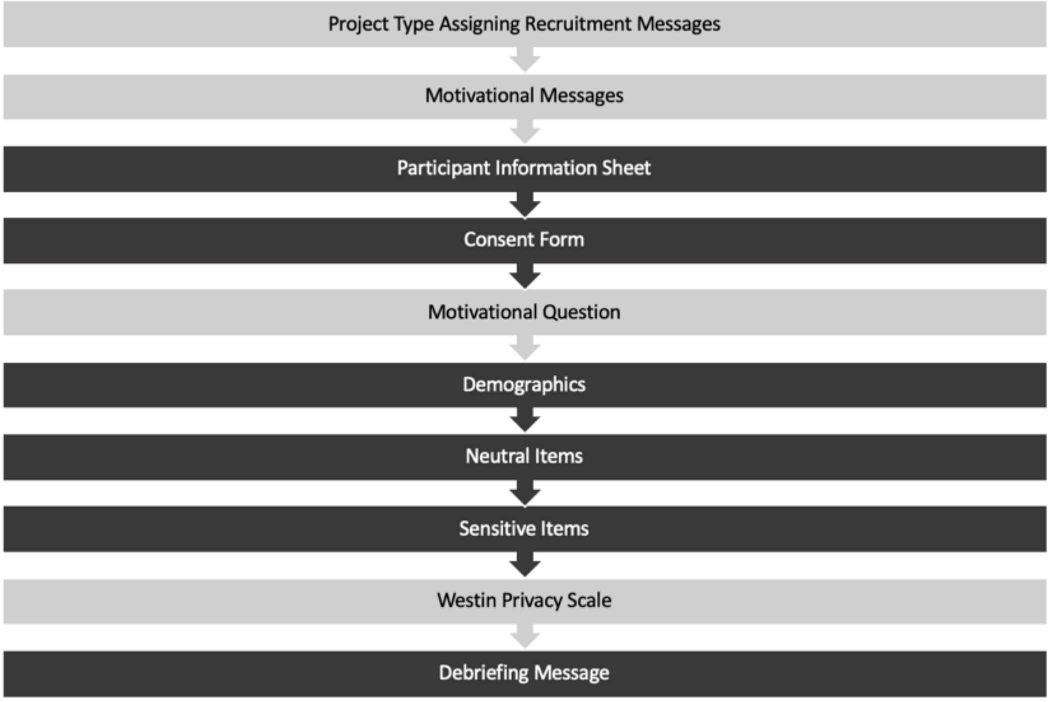


Figure 3 All survey elements

Some elements of the survey were present in each of the five studies: Participant Information Sheet, the Consent Form, the Demographics Questions, the Neutral Items, the Sensitive Items, and the Debriefing Message. In Figure 3, these elements are shaded in dark grey.

Other elements of the survey varied across the five studies. These were the elements that enabled us to operationalise the independent variables: project type assigning recruitment messages, motivational messages, a motivational question, and the Westin Privacy Scale. In Figure 3, these elements are shaded in light grey.

3.4.2.2. Motivational messages

In four studies (Studies 1-4), each participant was presented with a randomly assigned motivational message. These messages were employed to operationalise the independent variable of Motivation – a randomly assigned message prompted the participant to focus on a particular reward for taking part in citizen science. The respective content of messages used in Studies 1, 2, 3 and 4 will be discussed in the experimental chapters.

3.4.2.3. Participant information sheet and consent form

In each study, at the start of the survey, participants were presented with a Participant Information Sheet that outlined the project's purpose, participation requirements, data protection policy and a note outlining the possibility of the use of deception. This was followed by the main consent question (enquiring whether the participant wished to proceed with the study), a question enquiring about interest in future studies, and a question enquiring whether participants wished to be sent a copy of the research report upon publication, and fifteen secondary consent questions compliant with the General Data Protection Regulation. The Participant Information Sheets and the consent questions are included in Appendix A.

3.4.2.4. Demographics

In each study, participants were asked four questions about their demographics, so that we would be able to provide background information about the participant samples. These four questions inquired about the participants' age, gender, use of the Internet and previous experience with citizen science. The Demographics questions are included in Appendix A.

3.4.2.5. Motivational question

In Studies 2 and 5, participants were presented with a motivational question. The motivational question was employed to operationalise the independent variable of Preference – it enquired about participants’ preferences regarding the benefit of citizen science that should be prioritised in future projects. This variable is further discussed in the experimental chapters.

3.4.2.6. Neutral Items and Sensitive Items

In each of the five studies, participants were presented with a set of ‘Neutral Items’ and a set of ‘Sensitive Items’.

3.4.2.6.1. Neutral Items

The Neutral Items varied across the studies. In Studies 1-3 (sleep theme), the Neutral Items were adapted from the 19-item Morningness-Eveningness Questionnaire (Horne & Ostberg, 1976) and are listed in Appendix A. This included questions such as: ‘*Considering only your own ‘feeling best’ rhythm, at what time would you get up if you were entirely free to plan your day?*’, and ‘*How alert do you feel during the first half hour after having woken in the mornings?*’. The Morningness-Eveningness Questionnaire was chosen because it pertains only to people’s sleep habits and preferences and does not include items concerning health history nor any other types of sensitive data. Moreover, this questionnaire allows for a calculation of a personal score, which aided the comparison of two types of motivation to learn: learning about science and learning about oneself.

For Studies 4 and 5 (pet owner theme), we created a survey of 14 Neutral Items that enquired about a person’s pet, including questions such as ‘*Do you have more than one pet?*’, and ‘*Please tell us the exact age of your pet*’. These pet-related Neutral Items were designed such that the style of each question matched the style of the corresponding Sensitive Item, and they are listed in Appendix A.

3.4.2.6.2. Sensitive Items

The Sensitive Items remained the same across all five studies. They were drawn from a study of data disclosure in a credit card application context (Malheiros, Brostoff, et al., 2013).

Within this thesis, the phrase ‘sensitive data’ refers to information that citizen scientists might perceive as too private to disclose within a citizen science project. We sought to examine citizen scientists’ willingness to share not only the information that is legally classified as sensitive but also other types of data that people may wish to keep private such as those relating to their home, finances, or regarding third persons.

The Sensitive Items, listed in Appendix A, included questions such as ‘*Did any of your loved ones die while you were growing up?*’, and ‘*Do you give us permission to obtain a copy of your TV license payment history?*’. In the survey, the Sensitive Items were presented in a block titled ‘Sources of Stress’ (the Participant Information Sheet explained that the survey examined the links between stress and sleep or pet ownership, respectively).

3.4.2.6.3. Survey question validation study

We conducted a card sorting study to validate the questions used in the online surveys in the studies presented in this thesis. We wanted to test the assumption that the ‘Sensitive Item’ questions will be interpreted as requesting sensitive data, and that the ‘Neutral Item’ questions will be interpreted as requesting neutral data – by participants in a citizen science context. In other words, it was important for us to check that the information requests defined as sensitive in the survey are also viewed as sensitive by citizen science participants. Below, we report the validation study for neutral and sensitive survey items.

3.4.2.6.3.1. Participants

We recruited 12 participants to take part in a card sorting study. All participants were over the age of 18, and they were postgraduate students at the UCL Interaction Centre department.

3.4.2.6.3.2. Materials

The materials included in this validation study were the Participant Information Sheet (included in Appendix A) and two card sorting tasks hosted on the online platform Optimal Workshop (https://www.optimalworkshop.com). One task was titled '*Citizen science study about pets*' while the other task was titled '*Citizen science study about sleep*'.

3.4.2.6.3.3. Procedure

The study took place in research rooms within the department. Participants were tested separately. Each participant read a pen and paper Participant Information Sheet and gave written informed consent.

The study itself took place on the researcher's laptop computer, which allowed us to maintain full anonymity for the participants. The participants were told that they would be completing two card-sorting tasks. The order of tasks was counterbalanced, with 6 individuals beginning from the pet theme study, and 6 beginning from the sleep theme study.

Each task started with instructions presented on the screen. At the start of the sleep theme task, the participants saw the following instructions:

'*Citizen scientists are people who volunteer their time for research projects. One type of citizen science project is an online survey. Some surveys include questions that request sensitive data.*

Imagine that you are filling a citizen science survey that looks at the connection between sleep habits and stress. You will see a list of 33 questions and 4 *categories (Definitely Neutral, Somewhat Neutral, Somewhat Sensitive, Definitely Sensitive). Please assign the questions to categories.*

There are no right or wrong answers – we are interested in your opinions. For example, if you read a question, and think 'this question asks me to share neutral data', then please assign it to one of the neutral categories. If you read a question and think 'this question asks me to share sensitive data' then please assign it to one of the sensitive categories.'

In turn, at the start of the pet theme task, the participants saw the following instructions:

'*Citizen scientists are people who volunteer their time for research projects. One type of citizen science project is an online survey. Some surveys include questions that request sensitive data.*

Imagine that you are filling a citizen science survey that looks at the connection between pet ownership and stress. You will see a list of 28 questions and 4 categories (Definitely Neutral, Somewhat Neutral, Somewhat Sensitive, Definitely Sensitive). Please assign the questions to categories.

There are no right or wrong answers – we are interested in your opinions. For example, if you read a question, and think 'this question asks me to share neutral data', then please assign it to one of the neutral categories. If you read a question and think 'this question asks me to share sensitive data' then please assign it to one of the sensitive categories.'

Following instructions, the participants clicked an arrow and saw two lists of items. On the left, they saw a list of cards with Neutral and Sensitive Items presented in random order. On the right, they saw a list of 4 categories, which represented 4 levels of sensitivity: 'Definitely Neutral', 'Somewhat Neutral', 'Somewhat Sensitive', and 'Definitely Sensitive'.

The participants were shown a brief instruction message: '*Take a look at the list of items on the left. Please sort those items into the groups provided on the right.*' They then sorted the questions on the left into the categories on the right, until all questions were assigned to a category. They then clicked an arrow and saw a thank you message.

After completing both tasks, the participants were verbally debriefed.

3.4.2.6.3.4. Results

The survey validation study asked the participants to complete two card sorting tasks, where they assigned survey questions to four categories. All participants completed both a sleep-themed task and a pet-themed task.

The table below presents the popular placements matrix for the sleep theme card sorting task. This matrix shows what percentage of participants assigned a survey question to a particular sensitivity category.

Table 1 Popular placements matrix for the sleep-theme card sorting task

	Definitely Neutral	Somewhat Neutral	Somewhat Sensitive	Definitely Sensitive
Neutral Item questions				
If you went to bed at 11 p.m. at what level of tiredness would you be?	75%	25%		
At what time of the day do you think that you reach your 'feeling best' peak?	67%	8%	25%	
One night you have to remain awake between 4-6 a.m. in order to carry out a night watch. You have no commitments the next day, which ONE of the following alternatives will suit you best?	67%	33%		
When you have no commitments the next day, at what time do you go to bed compared to your usual bedtime?	67%	33%		
If there is a specific time at which you have to get up in the morning, to what extent are you dependent on being woken up by an alarm clock?	58%	42%		
You have to do two hours of hard physical work. You are entirely free to plan your day and considering only your own 'feeling best' rhythm which ONE of the following times would you choose?	58%	42%		
You wish to be at your peak performance for a test which you know is going to be mentally exhausting and lasting for two hours. You are entirely free to plan your day and considering only your own 'feeling best' rhythm which	58%	42%		

ONE of the four testing times would you choose?				
Assuming adequate environmental conditions, how easy do you find getting up in the mornings?	50%	42%		8%
Considering only your own 'feeling best' rhythm, at what time would you go to bed if you were entirely free to plan your evening?	50%	42%	8%	
For some reason you have gone to bed several hours later than usual, but there is no need to get up at any particular time the next morning. Which ONE of the following events are you most likely to experience?	50%	25%	25%	
How alert do you feel during the first half hour after having woken in the mornings?	50%	42%	8%	
One hears about 'morning' and 'evening' types of people. Which ONE of these types do you consider yourself to be?	50%	42%	8%	
Suppose that you can choose your own work hours. Assume that you worked a FIVE-hour day (including breaks) and that your job was interesting and paid by results. Which FIVE CONSECUTIVE HOURS would you select?	50%	50%		
How is your appetite during the first half-hour after having woken in the mornings?	42%	33%	25%	
You have to engage in hard physical exercise. A friend suggests that you do this for one hour twice a week and the best time for him is between 10-11 p.m. Bearing in mind nothing else but your own 'feeling best' rhythm how well do you think you would perform?	42%	42%	17%	
During the first half-hour after having woken in the morning, how tired do you feel?	33%	67%		
At what time in the evening do you feel tired and as a result in need of sleep?	42%	58%		
Considering only your own 'feeling best' rhythm, at what time would you get up if you were entirely free to plan your day?	42%	50%	8%	
You have decided to engage in some physical exercise. A friend suggests that you do this one hour a week and the best time for him is between 7:00-8:00 a.m. Bearing in mind nothing else but your own 'feeling best' rhythm, how do you think you would perform?	42%	50%	8%	
Sensitive Item questions				
What is the job of your partner / spouse? Please describe…		17%	58%	25%
Did you live with both your mother and father while you were growing up?	8%	17%	50%	25%
What is the length of the longest relationship you have had with a partner / spouse? (years/ months/ weeks)		8%	50%	42%
Could you list the names and either phone numbers or email addresses of three of your closest friends?				100%
May we obtain a copy of your insurance claims (e.g. car, house)?			8%	92%
Do you give us permission to contact your local council to get a copy of your council tax payment history?		8%	17%	75%
Please provide the name and address (or other contact details) of a previous employer so that we can request a copy of the last recommendation from him/her about you…		8%	17%	75%

What are the names of 3 people that you are friends with on a social networking site (facebook, twitter) whose profiles you would be happy to share with us? Please list…	8%	17%		75%
Did any of your loved ones die while you were growing up? Please give their relation to you (e.g. mother, brother, friend, etc.)			33%	67%
Do you give us permission to obtain a copy of your gas or electricity payment history?		8%	25%	67%
Do you suffer from any medical conditions? Please list...			33%	67%
Do you give us permission to obtain a copy of your TV licence payment history?		8%	33%	58%
Will you allow us to measure the typical number and length of messages between you and your friends on social networking sites?		8%	33%	58%
What are the names of 3 people that you are friends with on a professional networking site (LinkedIn, Orkut) whose profiles you would be happy to share with us? Please list...	8%	8%	42%	42%

The table below presents the popular placements matrix for the pet theme card sorting task.

Table 2 Popular placements matrix for the pet-theme card sorting task

	Definitely Neutral	Somewhat Neutral	Somewhat Sensitive	Definitely Sensitive
Neutral Item questions				
Does your pet ever wear a body warmer or any other type of pet clothing?	100%			
Are you planning to have more pets in the future?	83%	8%	8%	
Do you have more than one pet?	83%	17%		
Does the pet you currently own live with you?	83%	17%		
How many hours does your pet spend napping during the day?	83%	17%		
What are the 3 words that most accurately describe your pet? Please list...	83%	8%		8%
Could you list the 3 foods your pet enjoys most?	75%	25%		
Please tell us the exact age of your pet.	75%	8%	8%	8%
What is your pet's favourite toy? Please describe...	75%	17%	8%	
Did you own a pet when you were a child? Please specify the species of your childhood pet.	67%	17%	17%	
Do you buy birthday gifts for your pet?	58%	25%	17%	
Do you think that owning a pet nowadays is expensive?	58%	25%	17%	
What are the 3 things that most annoy your pet? Please list...	58%	33%	8%	
Does your pet have any health issues? Please list...	17%	33%	25%	25%
Sensitive Item Questions				
Did you live with both your mother and father while you were growing up?	8%	8%	58%	25%

What is the job of your partner / spouse? Please describe…		17%	58%	25%
Could you list the names and either phone numbers or email addresses of three of your closest friends?				100%
Do you give us permission to contact your local council to get a copy of your council tax payment history?		8%		92%
May we obtain a copy of your insurance claims (e.g. car, house)?		8%		92%
Do you suffer from any medical conditions? Please list...			17%	83%
What is the length of the longest relationship you have had with a partner / spouse? (years/ months/ weeks)		8%	17%	75%
Did any of your loved ones die while you were growing up? Please give their relation to you (e.g. mother, brother, friend, etc.)		8%	25%	67%
Do you give us permission to obtain a copy of your TV licence payment history?		8%	25%	67%
Do you give us permission to obtain a copy of your gas or electricity payment history?		8%	25%	67%
What are the names of 3 people that you are friends with on a social networking site (facebook, twitter) whose profiles you would be happy to share with us? Please list…	8%	8%	17%	67%
Please provide the name and address (or other contact details) of a previous employer so that we can request a copy of the last recommendation from him/her about you…			42%	58%
Will you allow us to measure the typical number and length of messages between you and your friends on social networking sites?		8%	33%	58%
What are the names of 3 people that you are friends with on a professional networking site (LinkedIn, Orkut) whose profiles you would be happy share with us? Please list...	8%	8%	42%	42%

As demonstrated in the two tables presented above, in both card sorting tasks, all of the Neutral Item questions were identified by the largest number of participants as either ‘Somewhat Neutral’ or ‘Definitely Neutral’. Similarly, all of the Sensitive Item questions were identified by the largest number of participants as either ‘Somewhat Sensitive’ or ‘Definitely Sensitive’. These findings confirm that the survey questions were accurately identified as requesting either sensitive or neutral data, by participants in a citizen science context – therefore suggesting that these questions are appropriate for the study of neutral and sensitive data among citizen scientists.

3.4.2.7. Westin Privacy Scale

In the case of Study 1, we looked at people's attitudes towards data privacy. Therefore, participants in Study 1 were presented with the 3-question Westin Privacy Scale (Harris and Associates Inc. & Westin, 1998), which required the participants to judge the truthfulness of three statements. This scale is further described in the experimental chapter.

3.4.2.8. Debriefing Message

In each study, the participants were presented with a debriefing message at the end of the survey. In Studies 2-5 we additionally sent the debriefing message via email to the participants who disclosed their email address. The debriefing message explained the aims of the study and reminded that, should the participants change their minds, they could withdraw from the study without having to provide a reason. Moreover, in each study, the Participant Information Sheet warned the participants that many psychological studies use deception and that the full aim of the study would be explained following the completion of the survey. Participants only became aware that they were taking part in a data disclosure study once they had completed the survey – specifically at the point of reading the debriefing message. This was necessary to ensure that we study participants' authentic behaviour as opposed to their opinions or attitudes.

3.4.3. Research Design

Studies presented in this thesis employ a strictly quantitative methodology. The choice of quantitative methodology serves the aim of achieving insight into how small changes in the design of citizen science projects might impact the degree to which participants disclose or withhold personal data. To our understanding, this systematic quantitative exploration of data disclosure in citizen science has not been undertaken previously.

Furthermore, it should be noted that ethical considerations influenced the choice of research design: to support transparency in the practice of citizen science, we steered away from exploring methods of extracting data from participants through non-overt means. The methodology of studying data disclosure behaviour allows researchers to establish a good understanding of how different elements of survey design (such as the order in which survey questions are presented) can act to extract larger volumes of data. However, we aimed to focus solely on studying the relationship between overt motivational cues and data disclosure, otherwise maintaining a transparent survey design.

The design of each of the five studies reported in this thesis is summarised in Table 3.

Table 3 Experimental Design in Studies 1-5

	Study Design	***Independent variables***
Study 1	4 x 3 independent ANOVA design	**Motivation** (Learning/Social/ Contribution/Altruism) **Attitude** (Fundamentalists/ Pragmatists/ Unconcerned)
Study 2	3 x 4 independent ANOVA design	**Motivation** (Learning about science/ Learning about self/Control) **Preference** (Learning/Social/ Contribution/Altruism)
Study 3	2 x 3 independent ANOVA design	**Project Type** (Citizen science/Online quiz) **Motivation** (Learning about science/ Learning about self/Control)
Study 4	Independent groups design	**Motivation** (Learning/Social/ Contribution/Altruism)
Study 5	Independent groups design	**Preference** (Learning/Social/ Contribution/Altruism)

3.4.3.1. Independent Variables

The independent variables studied in this thesis are Motivation (operationalised by presentation of motivational messages at the start of the survey), Attitude (operationalised by asking participants to fill a three-question Westin Privacy Scale), Preference (operationalised by asking participants to choose which benefit of citizen science should be prioritised in future projects), and Project Type (operationalised by presenting two variants of Twitter recruitment messages). Individual study reports include detailed discussions of each of the independent variables.

3.4.3.2. Dependent Variable

The dependent variable is this thesis is Disclosure. All studies in this thesis make use of the same dependent variable, which we will describe here.

The dependent variable was operationalised by asking participants to answer Neutral Item and Sensitive Item questions. This resulted in a disclosure score that was the number of items disclosed by the participant.

In the case of Neutral Items, the participant's disclosure score could range from 0 to 19 in Studies 1, 2, and 3 (where the role of Neutral Items was played by a 19-item questionnaire), and it could range from 0 to 14 in Studies 4 and 5 (where the role of Neutral Items was played by a 14-item questionnaire). **While we tracked and reported the volume of neutral data disclosed by the participants across all five studies, the main point of interest was the volume of disclosed sensitive data and the impact of independent variables on it**. In the case of Sensitive Items, the maximum number of disclosed items was 14 and so, the participant's disclosure score could range from 0 to 14. For each Sensitive Item question where participants disclosed at least one data item or agreed to their data being accessed, they were assigned one point. Where a participant skipped a question or failed to provide necessary details (for example stating that they suffer from a medical condition but not naming the condition), this was treated as withholding of data, and 0 points were assigned for the particular question item/s. Where a participant provided reasonable alternative data (for example stating that they do not have a partner in place of naming a partner's occupation), this was treated as disclosure and a point was assigned for the question item. In sum, the sensitive data disclosure score was the total of points assigned across all Sensitive Items.

3.4.4. Procedure

To emulate the authentic context of citizen science, for each study, we created a citizen science project that was advertised via Twitter. In each study, participants were recruited through Tweets that advertised the project and included a link to the survey (hosted on the Qualtrics platform). Only in Study 3, Twitter recruitment assigned participants to Project Type; in other studies, Twitter messages did not play any role other than the recruitment of participants.

After being redirected to the Qualtrics website but before starting the survey, participants in all studies but Study 5 were presented with one of the randomly assigned motivational messages.

Before filling the survey, all participants were instructed to read the Participant Information Sheet, which informed them about the aims of the study, participation requirements, possible risks and benefits of taking part, data management practices, use of deception, and their right to withdraw from the study. They were also informed that, in line with the estimation of the Qualtrics platform, it would take around 20 minutes to complete the survey. Participants were then asked to answer the main consent question (enquiring about whether they wished to proceed with the study after reading the information sheet), 2 optional questions about future contact and a request to receive a copy of the research publication, as well as 15 compulsory secondary consent questions.

Following consent, participants in Studies 2 and 5 were presented with a motivational question that asked them to choose one of the four benefits of citizen science that they would like to see prioritised in future projects. Following the motivational question (and directly after consent form in case of participants in Studies 1, 3, and 4), participants proceeded to answer 4 Demographics questions, that enquired about age, gender, frequency of Internet use, and whether a participant had ever taken part in a citizen science project.

Finally, participants were asked to answer a set of Neutral Item questions, followed by a set of Sensitive Item questions.

Only in Study 1, the survey ended with a 3-question Westin Privacy Scale.

At the end of the survey, all participants were redirected to a debriefing message, and in the case of individuals taking part in Studies 2 and 3, they were also shown their sleep score.

3.5. Limitations

The methodological approach chosen for this thesis had several limitations. Firstly, we focused solely on the format of an online survey. Therefore, our findings may not apply to other types of citizen science activities, such as providing ecological observations (for example, time- and location-stamped photos or written descriptions that include time, date and location) or online tests (such as those used in psychological citizen science, for example, to assess reaction speed). More specifically, by studying data disclosure in online citizen science surveys, we enquired into the overt disclosure among citizen scientists, and not some of the more pernicious types of data disclosure that people may not always be aware of – such as metadata or data that may not appear sensitive, however, can render a person vulnerable to manipulation (an example of that would be the personality tests that led to targeted political advertisements in the Cambridge Analytica scandal (Isaak & Hanna, 2018)). These are important issues and should be explored in further research. In the case of this thesis, the sole focus on one format of data disclosure in citizen science was seen as the optimal choice, due to the lack of previous experimental studies in this area. This narrow focus allowed us to establish the boundaries of the investigation into data disclosure in one-off citizen science projects and let us help build a foundation for future research on this topic.

Secondly, it would have been informative to conduct follow-up interviews with the participants who took part in our studies, to learn more about how people perceived data requests and how they made decisions about whether to share or withhold information. This approach was not followed for two reasons. Firstly, time restrictions and the difficulties associated with the recruitment of unpaid volunteers would have made it challenging to expand on the studies. Moreover, the main focus of this thesis was to provide experimental evidence for whether or not citizen scientists' disclosure is affected by what they are told about the benefits of participation. As such we were primarily interested in participants' spontaneous real-time disclosure decisions. Nevertheless, future research could benefit from a deeper understanding of how data disclosure in one-off citizen science projects is understood, negotiated and conceptualised by citizen scientists, and this could be achieved, for example, by combining experimental research with semi-structured interviews.

Furthermore, the recruitment strategy adopted for studies in this thesis presented some limitations. We focused on recruitment on the social platform Twitter. As a result, we may have narrowed the population of participants, which could be more diverse if we had expanded recruitment to other social platforms such as Instagram, Facebook or discussion boards. On the other hand, managing several social media accounts and spreading the available advertisement funds over several platforms could have made recruitment more complex and diverted the attention away from completing the studies within time constraints. Alongside recruitment of a more diverse sample, it would have been interesting to study how different approaches to recruitment may affect the data disclosure behaviour of the resultant samples (for example, comparing whether posting a logo of the university next to the survey link could increase trust and therefore result in a larger volume of disclosure). These issues would benefit from being studied in future research. In the case of this thesis, the focus remained on clarifying the relationship between communication about the benefits of citizen science and disclosure behaviour.

Finally, there were limitations present regarding the Materials chosen for our studies. Two questions in the Sensitive Items (adapted from (Malheiros, Brostoff, et al., 2013)) alluded to topics specific to respondents resident in the UK: one relating to council tax and another to the TV licence payment. We encouraged participation from an international sample, and we are aware that local government structure and TV licencing systems vary across different countries and regions. Nevertheless, all questions were formulated using clear language and, in the case of confusion, participants could easily contact the researcher via social media to seek clarification. We, therefore, decided not to alter these questions as that would have interfered with our goal of accurate replication of the original study from which the Sensitive Items were adapted. Another limiting factor was that the 14 Sensitive Item questions only allowed us to draw limited conclusions about the disclosure behaviour. In the future, it will be important to study a broader range of sensitive data requests, to provide insight into how different levels of information sensitivity may impact disclosure decisions in a citizen science setting.

3.6. Conclusion

This chapter provided an overview of the method employed in this thesis. Starting from a rationale for the choice of contribution format (online survey), and a justification for studying the relationship between incentives and disclosure, we described the recruitment, materials, research design and procedure used in this thesis. We also discussed the limitations of the current approach and how our approach can be developed in future studies.

The following chapter reports the first three of the five studies presented in this thesis. These three studies were conducted by asking participants to fill an online survey on the theme of sleep habits. They explored the impact of the independent variables of *Motivation*, *Attitude*, *Preference*, and *Project Type*, on the volume of data disclosed by citizen scientists.

Chapter 4

Motivation to learn: Impact on disclosure of sensitive data among citizen scientists

4.1. Outline

This chapter describes three studies that looked at data disclosure in a citizen science context. In each study, participants were presented with a randomly assigned motivational message, before proceeding to answer a set of questions that require disclosure of neutral (non-sensitive) data followed by a set of questions that require the disclosure of sensitive data. All three studies aimed to tease out what incentives best motivate citizen scientists to disclose sensitive data. Participants' susceptibility to motivational cues in these studies provides empirical evidence for the importance of transparency in the communication about the benefits of citizen science.

Study 1 utilises four motivational messages that emphasise the social, contribution, altruistic or learning benefits of citizen science. Study 2 compares two different facets of the motivation to learn: learning about science and learning about oneself – as well as contrasting the impact of motivational messages and a motivational question. Study 3 looks at the role of motivational messages across two different types of projects: a citizen science survey and an online quiz.

The three studies reported in this chapter demonstrate that citizen scientists recruited for a one-off survey are far less likely to share sensitive, as opposed to neutral, information. While difficult to obtain, disclosure of sensitive data in citizen science can be increased by the presentation of motivational messages that emphasise learning at the start of a project. Moreover, it is important to highlight the citizen science nature of a project at the recruitment stage, as individuals who sign up for a citizen science project tend to disclose a larger volume of sensitive data when compared to individuals who sign up for an online quiz – even when the content of both projects is identical.

4.2. Study 1 *Sleep Mapping*: Selective disclosure of sensitive data among citizen scientists

This study was published as Rudnicka, Cox, and Gould (2013) and presented at the annual ACM CHI Conference on Human Factors in Computing Systems.

4.2.1. Introduction

We describe the first study to experimentally investigate the disclosure of personal data in an authentic citizen science scenario. This study aims to identify the reasons behind citizen scientists disclosing or withholding personal data when participating in citizen science initiatives. To simulate an authentic citizen science context, we created a fictional citizen science project called *Sleep Mapping*, in which participants were randomly divided into four motivational groups and answered both neutral as well as sensitive requests for personal data. They also answered three questions that make up the Westin Privacy Scale. We anticipate that the type of motivational messages presented to a participant will have an impact on the volume of sensitive data that they disclose. Moreover, we expect that *Privacy Fundamentalists*, as classified on the Westin Privacy Scale, will disclose a smaller volume of sensitive data than *Privacy Pragmatists* and *Privacy Unconcerned*.

The current study sets out to explore the impact of the variables of Motivation and Attitude on the volume of sensitive data disclosed by the participants.

4.2.2. Independent Variables in Study 1

Figure 4 depicts the elements of the Sleep Mapping survey.

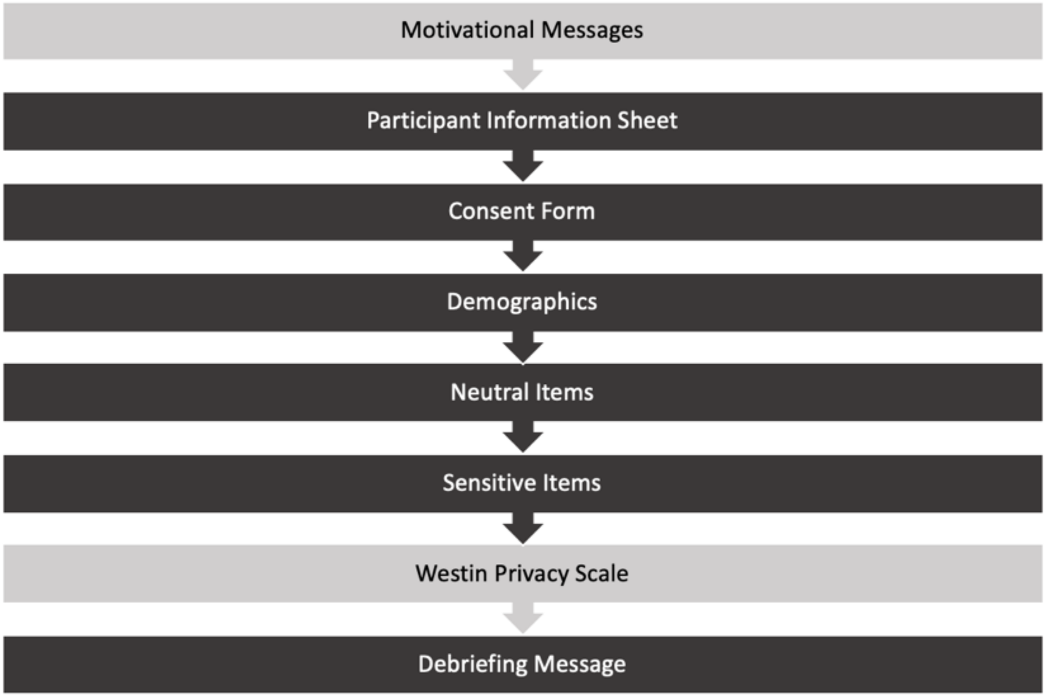


Figure 4 Elements of the Sleep Mapping survey

In Study 1, the independent variables were Motivation and Attitude. Motivation had 4 levels (Learning, Social Proof, Contribution, and Altruism). It was operationalised by presenting the participants with one of the four randomly assigned motivational messages: '*Extend your knowledge of health psychology by participating in the Sleep Mapping survey!*' (Learning); '*Join your fellow citizen scientists in establishing connections between stress and sleep!... Many citizen scientists are already participating in the project…*' (Social Proof); '*You can contribute to science by answering questions about sources of stress in your life and quality of sleep!*' (Contribution); '*Health psychology needs your help to connect sources of stress to patterns of sleep behaviour!*' (Altruism).

The variable of Attitude had 3 levels (Fundamentalists, Pragmatists, Unconcerned). It was operationalised by asking participants to fill the 3-question Westin Privacy Scale (Harris and Associates Inc. & Westin, 1998), 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 *Privacy Fundamentalists* who are defensive about sharing personal data (agreed with the first statement and disagreed with the second and third statements), or *Privacy Unconcerned* who do not have major concerns about parting with their 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 (Harris and Associates Inc. & Westin, 1998).

4.2.3. Results

4.2.3.1. Participant attrition

A total of 331 individuals decided to take part in the survey, by agreeing to the main consent question. The attrition of participants throughout the survey is illustrated in Figure 5.

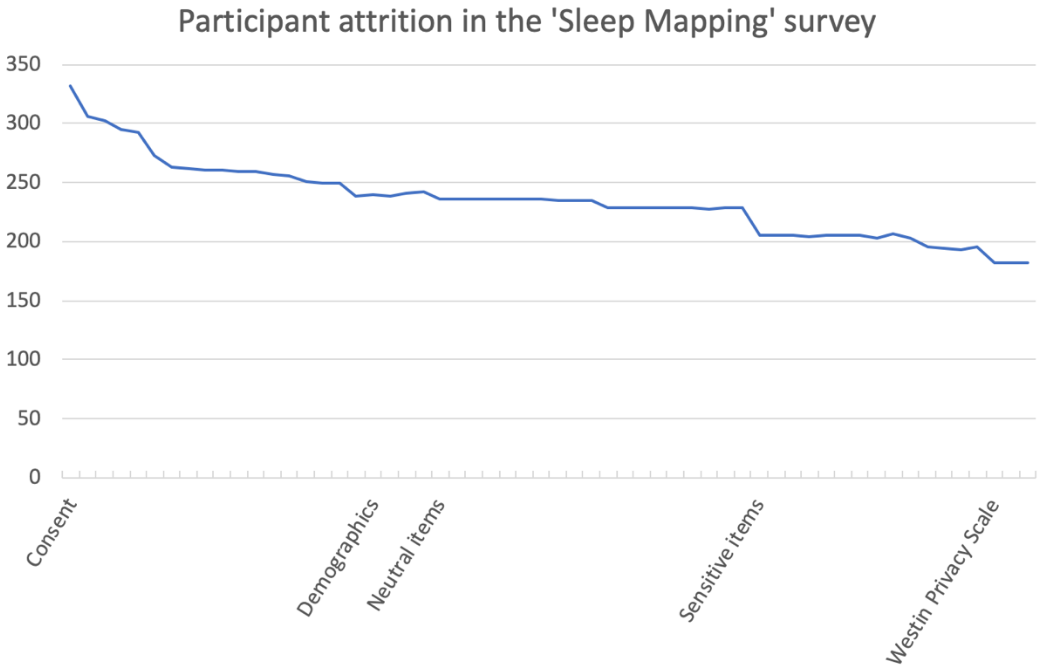


Figure 5 Participant attrition across survey stages in Study 1

Attrition over the 17 secondary consent questions resulted in a sample of n = 240 at the beginning of Demographics items. Twelve participants dropped out across these items. A decline in participant numbers was then observed at the point of the first question requesting sensitive data: 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 Items in the survey than the cumulative impact of questions requesting sensitive data, for participant attrition.

A final decline can be observed when progressing from the last question requesting sensitive data, to the Westin privacy scale: here, 13 participants dropped out before the Westin Scale. In total, a sample of n = 182 completed the last item of the survey.

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, 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).

4.2.3.2. Demographics

The final sample (defined as having reached and answered the last question in the survey) consisted of 182 participants.

Data about age, as reported by the participants, were available for n = 181, with ages ranging from 19 to 71 years old (mean = 37, SD = 11).

Data about gender, as reported by the participants, were as follows: n = 128 female, n = 53 male, and 1 participant who did not disclose their gender.

Data about Internet usage were available for 181 participants: n = 115 reported using the Internet ‘all the time’, n = 61 reported using the Internet ‘several times per day’, and n = 5 reported using 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.

More than half of the 182 participants reported having no previous experience of citizen science participation (n = 115), while 21 participants were ‘not sure’ about whether they had participated previously or not. Only 46 reported having taken part in a citizen science project in the past (n = 20 once before, and n = 26 several times).

4.2.3.3. Data disclosure across Neutral Items

Disclosure on the 19 Neutral Items was very high, ranging from 17 to 19 disclosed items (mean = 18.98, SD = 0.17), equivalent to 99.9% disclosure of neutral data.

4.2.3.4. Data disclosure across Sensitive Items

Disclosure on the 14 Sensitive Items was relatively low. It ranged from 3 disclosed items to 12 disclosed items (mean = 5.38, SD = 2.12), equivalent to 38.4% disclosure of sensitive data. As the counts were uneven in this study (19 Neutral Items and 14 Sensitive Items), they were transformed into percentages before a comparison between the two was conducted. The result of a paired t-test (t = 54.901, df = 181, p = .000, 2-tailed) suggested that the disclosure of Neutral Items was significantly higher than the disclosure of Sensitive Items.

4.2.3.5. Motivation

We anticipated that the variable of Motivation may have an impact on the disclosure of sensitive data. Participants who saw a message emphasising the Learning motivation (n = 42) shared the largest volume of sensitive data (m = 6.31, SD = 2.94), followed by participants (n = 50) in the Altruism group (m = 5.26, SD = 1.76), and participants (n = 50) in the Contribution group (m = 5.02, SD = 1.89). The lowest disclosure (m = 5.00, SD = 1.47) was demonstrated by participants in the Social Proof group (n = 40). A one-way between-subjects Analysis of Variance showed that the impact of Motivation on sensitive data disclosure was significant $F(3,178) = 3.818$, $p = .011$, $\eta = .060$. A post hoc Tukey test showed that the Sensitive Item disclosure in the Learning group differed significantly from the Social Proof group ($p = .024$), the Contribution group ($p = .018$), but not from the Altruism group ($p = .077$). Moreover, the differences between Social Proof and Contribution, Social Proof and Altruism, and Altruism and Contribution groups, were not found to be statistically significant. These results suggest that the Learning message was solely responsible for the differences in Sensitive Item disclosure across the Motivation variable.

4.2.3.6. Attitude

We anticipated that *Privacy Fundamentalists* may be less likely to disclose information. Based on their answers on the 3-question 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 sensitive data disclosure (m = 5.54, SD = 2.03) than *Privacy Fundamentalists* (m = 5.29, SD = 2.31) and *Privacy Unconcerned* (m = 4.45, SD = 1.44). However, a one-way between-subjects Analysis of Variance showed that the impact of the Attitude variable on the disclosure of Sensitive Items was not statistically significant ($F(2,179) = 1.393$, $p = .251$, $\eta = .015$).

4.2.3.7. Interaction between Motivation and Attitude

There was no interaction effect between the variables of Motivation and Attitude ($F(6,170) = .283$, $p = .944$, $\eta2 = .01$), suggesting that motivational messages can impact people’s willingness to share sensitive data, regardless of their attitudes towards data privacy. It also suggests that the type of motivational message a participant viewed did not have an impact on their performance on the Westin Privacy Scale.

4.2.4. Discussion

4.2.4.1. Data Disclosure

This study investigated the degree to which motivational messages and privacy attitudes influence the disclosure of sensitive data in a citizen science setting. As anticipated, we found that participants were far more likely to disclose data when faced with questions requesting neutral information, than when answering questions about sensitive issues. Similar to Malheiros et al.’s participants

(Malheiros, Brostoff, et al., 2013) 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 alertness regarding the types of disclosure requests they face.

4.2.4.2. Motivation

In this study, the variable of Motivation had a significant impact on the disclosure of sensitive data. Participants primed by the Learning motivational message were more likely to disclose sensitive data, than participants primed by other motivational messages. It appears that presenting a motivational message can lead to a participant disclosing a larger volume of data, even when the data requests pertain to sensitive personal information. This is in line with both the link between motivation and disclosure identified by Bowser et al. (Bowser et al., 2017) 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 (Crain, 2018). From a practical standpoint, our findings suggest that, when emphasised through short messages, certain types of motivations can result in a larger volume of disclosed data than others, and therefore better support the primary goals of projects that are focused on gathering personal data from citizen scientists.

It should, however, be investigated whether Learning is the main type of motivation that will encourage disclosure of data or whether sensitivity to motivational messages can vary across different samples. Future research should explore the stability of the motivation-disclosure relationship across different participant populations. Moreover, it is important to clarify whether motivational messages tap into, or indeed override, the preferences of citizen scientists.

Despite the fact that our findings reached statistical significance, it should be noted that the differences in mean Sensitive Item disclosure across motivational groups were relatively small (45.1% for participants primed by Learning, compared to 35.7% for participants primed by Social Proof, and 35.9% and 37.6% for those who were presented with Contribution and Altruism messages, respectively). To identify useful design implementations that could help increase data disclosure in citizen science projects, more research needs to be conducted, exploring the efficacy of various motivational devices across different citizen science samples.

4.2.4.3. Attitude

We found that *Privacy Fundamentalists*, as classified by answers to the 3-question Westin's Privacy Scale (Harris and Associates Inc. & Westin, 1998), were not significantly 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 (Norberg, Horne, & Horne, 2007). Individuals in Study 1 may have perceived the rewards of citizen science participation to be greater than the cost of disclosing personal data. This is in line with the results of Bowser et al. (Bowser et al., 2017) who found that participants decided to share information even when they had some privacy concerns because their concerns were overridden by the specific motivations that made them want to contribute to citizen science.

It should be noted that at this point we are not able to directly compare the results demonstrated by Malheiros et al. (Malheiros, Brostoff, et al., 2013) from whom we adapted the Sensitive Item questions. These researchers found that *Privacy*

Fundamentalists were less likely to submit the form in their study. In Study 1, the Westin scale was administered at the very end of the survey (and therefore only to participants who have not dropped out prior to that 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 spontaneous disclosure decisions earlier in the survey.

4.2.5. Limitations

This study explored only a limited range of sensitive and neutral data requests. Including a wider range of questions in future research could shed light on why particular citizen scientists decide to answer some, but not other, requests for sensitive data. 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. Reasons for participant attrition should be explored in future studies.

4.2.6. Conclusion

In this study, we demonstrated the existence of selective disclosure of data in a citizen science context. Participants were significantly more likely to disclose sensitive data when they were presented with a motivational message that emphasised learning opportunities, than when they saw a message emphasising contribution or social proof. To our knowledge, this is the first study to experimentally investigate antecedents of disclosing or withholding personal information by citizen scientists. Secondly, the findings of this study 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.

4.3. Study 2 *Sleep Experience*: Motivational messages and motivational preferences

4.3.1. Introduction

Study 1 demonstrated that when eliciting disclosure of sensitive data among citizen scientists, emphasising learning opportunities is significantly more effective than emphasising contribution or social factors. While we do know that different messages will result in different volumes of disclosed information, we do not know how a lack of any communication about the benefits of participation might impact disclosure. It is important to experimentally compare the presence and absence of motivational messages.

Furthermore, while the participants in Study 1 were encouraged to '*Extend your knowledge of health psychology by participating in the Sleep Mapping survey!*', it is unclear whether the effectiveness of this message stems from citizen scientists' general curiosity about the subject area, or a willingness to better understand their sleep patterns by participating in the survey. Although some publications have separated citizen science from the desire to learn something about oneself (for example, Wiseman and colleagues (Wiseman et al., 2017) see the latter as part of the quantified-self movement), a number of psychological citizen science projects use the incentive to learn about oneself to recruit participants. A recent study highlighted the interest of citizen scientists in learning about themselves. August et al. (August et al., 2018) demonstrated that, for some projects, participant recruitment could be increased by 9% by changing the framing from 'supporting science' to 'comparing oneself to others'. To aid the better design of future projects, it is valuable to explore the relative effectiveness of 'learn about science' and 'learn about oneself' messages, when appealing to citizen scientists for sensitive data.

In Study 1, we found that motivational messages emphasising learning opportunities, presented to participants at the start of the survey, were more effective at eliciting disclosure of sensitive data, than messages focused on contribution or social factors. However, we do not know whether such motivational messages gave participants a real-time opportunity to evaluate their preferences and weigh them against the cons of participation. In other words, did citizen scientists disclose a greater volume of sensitive data when presented with a message about learning, because this message aligned with their underlying preferences? Or did this motivational message briefly alter how participants perceived the value of filling the survey?

To better understand the way that motivational messages impact disclosure, it is important to explore whether participants’ motivational preferences might account for the degree to which they disclose or withhold data. In Study 1, the Learning group was the second smallest group out of four (due to participant attrition) – and yet its participants demonstrated the largest volume of sensitive data disclosure. Therefore, it is likely that the most popular motivational preferences will not be the ones associated with the highest data disclosure.

This study, named *Sleep Experience*, looks at the relative effectiveness of the two facets of motivation to learn: learning about science and learning about oneself. Disclosure of sensitive data is compared across three motivational groups: 1) emphasising learning about science, 2) emphasising learning about oneself, and 3) a control group presented with a message that welcomes people to the survey. We anticipate that these three messages will differentially impact the volume of data that citizen scientists disclose. Moreover, this study looks at people’s motivational preferences – by asking participants about which benefits of citizen science should be prioritised in future projects (a ‘motivational question’, separated from the motivational messages by the Participant

Information Sheet and consent form), we investigate the impact of motivational preferences on data disclosure.

This study sets out to test two hypotheses:

H1: Citizen scientists will disclose a larger volume of sensitive data when presented with a message emphasising learning about science than when presented with a neutral message or a message emphasising learning about oneself.

H2: Personal motivational preferences of citizen scientists will have an impact on the volume of sensitive data they disclose.

4.3.2. Independent Variables in Study 2

Figure 6 depicts the elements of the Sleep Experience survey.

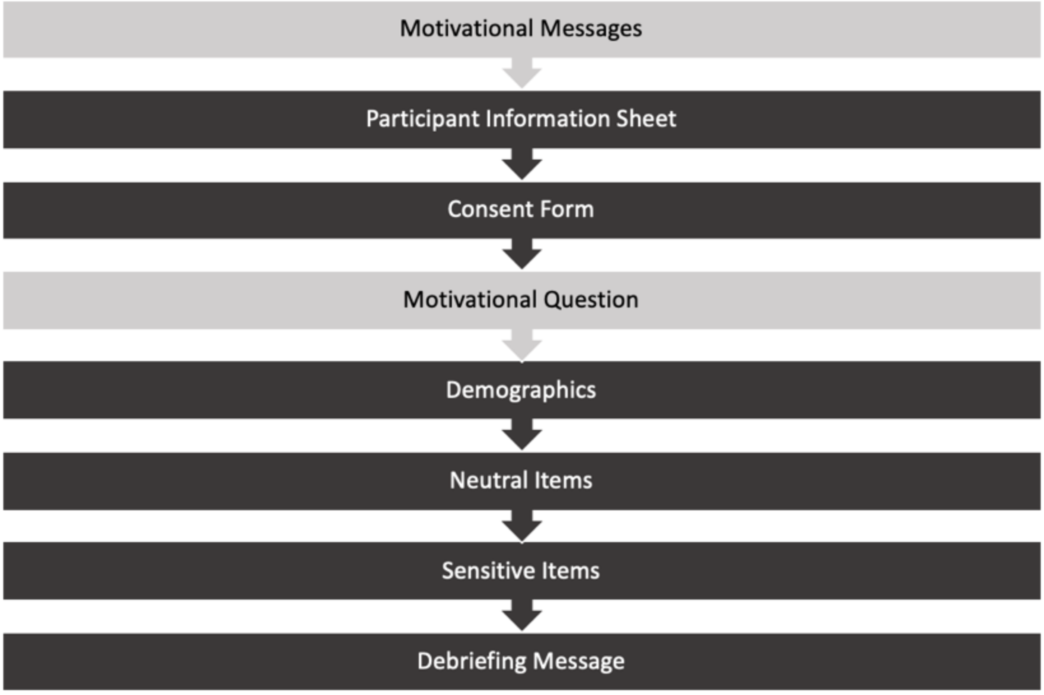


Figure 6 Elements of the Sleep Experience survey

In Study 2, the independent variables were Motivation and Preference. Motivation had 3 levels (Learning about science, Learning about self, Control). It was operationalised by presenting the participants with one of the three randomly assigned motivational messages: ‘*Extend your knowledge of health psychology by participating in the Sleep Experience survey!*’ (Learning about science), ‘*Fill the Sleep Experience survey and discover your sleep score!*’ (Learning about self); ‘*Welcome to the Sleep Experience survey!*’ (Control).

The variable of Preference had 4 levels (Learning, Altruism, Contribution, and Social interaction). It was operationalised by asking participants to answer a motivational question that stated: ‘*Before filling the survey, please answer this question to help us design better citizen science projects: Which benefit of taking part in citizen science is most important to you?*’. Participants were given four options from which to choose one: 1) Learning about science, 2) Social interaction, 3) Making a contribution to a project, 4) Helping scientists.

4.3.3. Results

4.3.3.1. Participant Attrition

A total of 571 individuals decided to take part in the survey, by agreeing to the main consent question. The attrition of participants throughout the survey is illustrated in Figure 7.

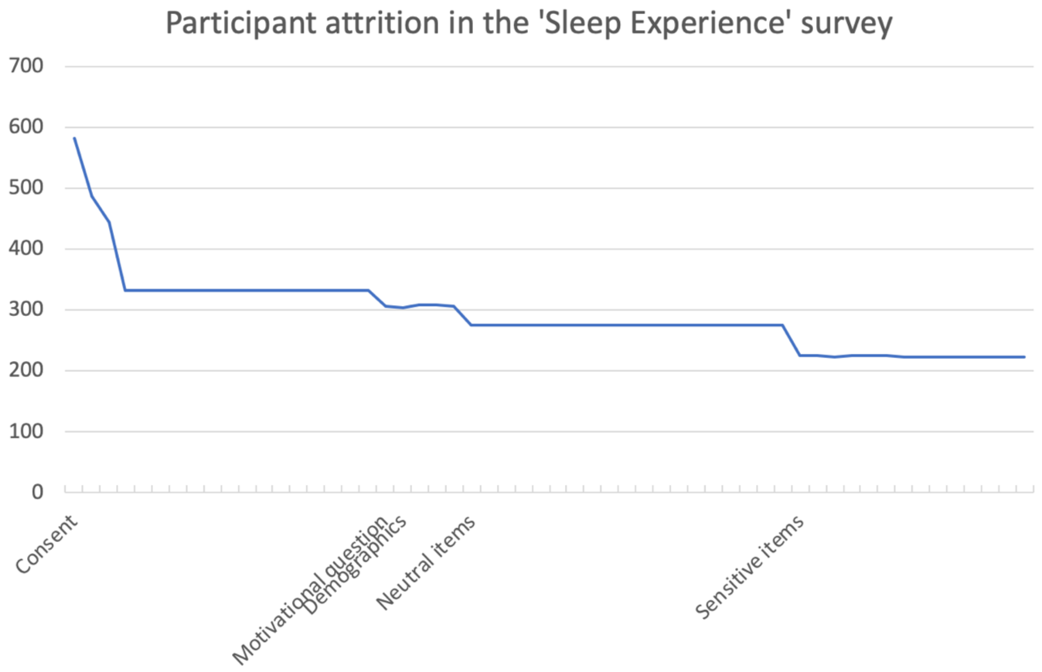


Figure 7 Participant attrition across survey stages in Study 2

Attrition over the 17 secondary consent questions resulted in a sample of n = 304 at the beginning of Demographics items. A sample of 275 answered the first question requesting neutral data, and no attrition occurred over the course of Neutral Items. A decline in participant numbers was then observed at the point of the first question requesting sensitive data: out of 275 participants who answered the last Neutral Item, only 224 answered the first Sensitive Item. This 18.5% 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 1 participant dropped out over the course of Sensitive Items. This suggests, similar to Study 1, a greater role of the point of change from neutral to sensitive questions in the survey, than the cumulative impact of questions requesting sensitive data, for participant attrition. In total, a sample of n = 223 completed the last item of the survey. One participant was then eliminated, as they had stated their age as 17 years old and, therefore, did not meet the recruitment requirements. This resulted in a final sample of n = 222.

4.3.3.2. Demographics

The final sample (defined as having reached and answered the last question in the survey) consisted of 222 participants.

Data about age, as reported by the participants, were available for n = 219, with ages ranging from 18 to 76 years old (mean = 33, SD = 13).

Data about Internet usage were available for all participants: n = 138 reported using the Internet ‘all the time’, n = 81 reported using the Internet ‘several times per week’, n = 1 reported using the Internet ‘most days’ and n = 2 reported using the Internet ‘several times per week’.

Data about gender, as reported by participants (available for all participants) were as follows: n = 84 male, n = 129 female, n = 5 non-binary, n = 3 preferred not to answer, and n = 1 answered in their own words.

Only a small proportion of participants reported having had previous experience with citizen science (n =13, ‘once’, and n = 11, more than once’), while the majority reported either not having any citizen science experience (n = 165) or being ‘not sure’ (n = 33).

4.3.3.3. Data disclosure across Neutral Items

All participants answered every one of the 19 Neutral Item questions, resulting in 100% disclosure of Neutral Items in this study.

4.3.3.4. Data disclosure across Sensitive Items

Disclosure of sensitive data in this study ranged from 2 to 13 disclosed items (out of 14), with a mean of 4.84 items (SD = 2.05), equivalent to 34.6% disclosure of sensitive data. There was a small positive correlation of data disclosure with the age of participants (r = .195, n = 219, .004), suggesting that older citizen scientists are more likely to share a greater volume of sensitive data. As the counts were uneven in this study (19 Neutral Items and 14 Sensitive Items), they were transformed into percentages before a comparison between the two was conducted. The result of a paired t-test (t = 66.685, df = 221, p = .000, 2-tailed) suggested that the disclosure of Neutral Items was significantly higher than the disclosure of Sensitive Items.

4.3.3.5. Hypothesis 1: Motivation

Although individuals who followed the link to the survey were randomly assigned to three Motivation groups in equal numbers, due to the pattern of attrition in this sample, the final sample (n = 222) included slightly uneven numbers of participants in the 3 motivational groups: the ‘Learn about science’ group (n = 63), the ‘Learn about self’ group (n = 82), and the ‘Control’ group (n = 77).

Mean disclosure was 5.11 items (SD = 2.42) for participants in the ‘Learn about science’ group, m = 4.44 items (SD = 1.47) for participants in the ‘Learn about self’ group, and m = 5.04 items (2.20) for participants in the ‘Control’ group.

A One-Way Independent Analysis of Variance was not significant F(2,219) = 2.523, p = .083, η = .023, suggesting a lack of the impact of the Motivation variable on data disclosure. Hypothesis 1 was rejected.

4.3.3.6. Hypothesis 2: Preference

Across the four Preference groups (to which participants self-assigned by stating which benefits of citizen science ought to be prioritised in future projects), data were available for all but one participant who did not answer the motivational question. The mean disclosure was 4.81 items (SD = 2.09) in the Learning group (n = 48), 5.29 items (SD = 2.64) in the Social group (n = 17), 4.65 items (SD = 2.00) in the Contribution group (n = 97), and 5.00 items (SD = 1.91) in the Altruism group (n = 59). The differences in disclosure between groups were not found to be significant (F(3,217) = .677, p = .567, η = .009), suggesting a lack of a relationship between participants’ preferred benefits influence of citizen science, and how much sensitive data they share in citizen science projects. Hypothesis 2 was rejected.

4.3.3.7. Interaction between Motivation and Preference

There was no interaction effect between the variables of Motivation and Preference ($F(6,209) = .774$, $p = .591$, $\eta2 = .022$), suggesting that motivational messages can impact people's willingness to share sensitive data, regardless of which benefits of citizen science they prioritise. It also suggests that the type of motivational message a participant viewed did not have an impact on the answer they gave to the motivational question.

4.3.4. Discussion

4.3.4.1. Hypothesis 1: Motivation

We set out to explore whether motivation to learn about science and motivation to learn about oneself will differentially influence the disclosure of data in a citizen science context. Not only did we not find significant differences between the two types of motivation but also neither of those messages prompted people to share a larger volume of data than a neutral welcome message. In conjunction with findings from Study 1, these results suggest that motivational messages about learning do not increase data disclosure in and of themselves but solely prompt higher disclosure than messages focused on social or contribution aspects of citizen science participation.

However, a different factor may be responsible for the current results. During the recruitment process, we ensured that the tweets with a link to the survey were phrased neutrally, with no references to specific motivations such as an opportunity to learn or social factors. In the process of recruiting participants for Study 2, we sent out several tweets such as '*take part in our citizen science survey about sleep*' or '*take our SLEEP survey*' – we noticed that the tweets which spoke solely about a sleep survey and made no mention of citizen science proved the most popular, attracting the highest number of likes, retweets, and link clicks. It is, therefore, possible that participants in the current study were less focused on the ‘citizen science’ aspect of the survey and more on the opportunity to take an online survey/quiz. This proposition is in line with the smaller proportion of people who had previous experience of citizen science in this sample (10.7% of the total number of participants) when compared to Study 1 (25.3% of the total number of participants). Therefore, it would be worthwhile to compare disclosure behaviour across two modes of participant recruitment: one that frames the survey as citizen science, and one that frames the survey as an online quiz.

4.3.4.2. Hypothesis 2: Preference

Participants’ preferences did not account for the pattern of data disclosure. These findings suggest that, in contrast to committed citizen scientists (Bowser et al., 2017), the motivations that make dabblers view citizen science participation as worthwhile are not the same motivations that cause them to share sensitive data.

Interestingly, the most popular benefit of citizen science, Contribution, chosen by 97 participants, was also associated with the smallest volume of data disclosure. In contrast, the benefit of Social interaction was the least popular among participants, however also associated with the highest data disclosure. It should be noted that in the case of this citizen science project, participants were not offered any opportunities for social interaction with other participants (other than being able to send and see tweets about participation). It is possible that the individual characteristics of people who care about social interaction are also associated with an increased willingness to share sensitive data.

4.3.5. Limitations

This study focused on the exploration of data disclosure in one-off citizen science projects and therefore the finding that participants’ motivational preferences do not have an impact of the disclosure of sensitive data may not extend to all citizen science projects. In future research, it is important to explore which features of participation create a well-developed awareness of motivational preferences among citizen scientists and how these preferences may be channelled into a positive attitude towards data disclosure within projects.

4.3.6. Conclusions

In this study, neither randomly assigned motivational messages, nor participants' preferences, influenced the degree to which citizen scientists disclosed or withheld sensitive data. However, these results may have been affected by the lack of mention of 'citizen science' in the most popular recruitment Tweet. To clarify this, it is important to study whether the emphasis on the concept of 'citizen science' at the stage of recruitment may influence later disclosure behaviour within the survey.

4.4. Study 3 *Sleep Patterns*: Citizen science versus online quiz

4.4.1. Introduction

Study 1 demonstrated that short motivational messages can have an impact on the volume of sensitive data that citizen scientists decide to disclose: we showed that exposing citizen scientists to a message that emphasised the opportunity to learn about psychology, was associated with higher levels of data disclosure than in the case of messages emphasising social and contribution opportunities. In Study 2 we attempted to further explore the motivation to learn by comparing how the disclosure of sensitive data may be influenced by two types of learning: learning about science and learning about oneself. However, neither the motivational messages nor participants' motivational preferences could account for the pattern of data disclosure in Study 2.

It is important to note that the mean data disclosure in Study 2 (m = 4.84) was noticeably lower than in Study 1 (m = 5.38). Two factors could explain this. Firstly, the inclusion of a motivational question might lower participants' willingness to share data, by making them examine more critically the benefits of taking part in the study. Secondly, it is possible that the way the participants were recruited, influenced the results of Study 2. Namely, we noticed that a Tweet which encouraged participation in a sleep survey but did not mention citizen science, received the most likes, retweets and clicks. It is therefore possible that, when joining the study, some participants perceived it as an online survey unrelated to academic research. Comparing the disclosure behaviour of participants in a citizen science survey to disclosure behaviour in a survey advertised as an online quiz could shed more light on the results of Study 2.

Indeed, it is important to establish the extent to which disclosure of information among citizen scientists may differ from the disclosure of data in other online contexts. Traditionally, a distinction has been made between citizen science projects, which offer opportunities to learn about science or the scientific process, and projects that attract participants due to people’s desire to learn something about themselves (Wiseman et al., 2017). In recent years, however, the latter incentive has become a popular method of attracting participants to citizen psych-science (citizen science projects in the field of psychology (Jennett et al., 2014)).

Study 3 seeks to examine the differences between the way people disclose data when they join a citizen science survey versus when they join an online quiz that allows them to learn something about themselves. By varying both the types of tweets used to recruit participants, as well as the types of messages presented to those who click the survey link, this study will examine people’s sensitivity to motivational cues at different stages of participation.

We set out to compare data disclosure behaviour between two online contexts: 1) a citizen science survey and 2) an online quiz that allows people to find out their ‘sleep score’. Additionally, in both contexts, we will explore the impact of motivational messages evoking two types of learning motivations: learning about science (more aligned with citizen science projects) and learning about oneself (more aligned with online quizzes). In effect, we will be able to compare the relative impact – on data disclosure – of information presented at the point of recruitment versus the impact of the messages presented at the beginning of the survey. In other words, we will gain insight into the relative importance of recruiting a particular sample of participants versus adequately motivating already recruited participants – on the ability of citizen science projects to elicit disclosure of the sensitive data necessary for their successful completion.

We predict that participants in the ‘citizen science’ group will share a significantly larger volume of sensitive data than the participants in the ‘online quiz’ group. Moreover, we expect that the relationship between motivational message and the volume of sensitive data disclosed by the participants may be influenced by the type of project that these participants join in the first place.

The current study sets out to test two hypotheses:

H1: Participants who join a citizen science project will disclose a larger volume of sensitive data than participants who join an online quiz.

H2: The relationship between the motivational messages and the disclosure of sensitive data will be moderated by project type.

4.4.2. Independent Variables in Study 3

Figure 8 depicts the elements of the Sleep Patterns survey.

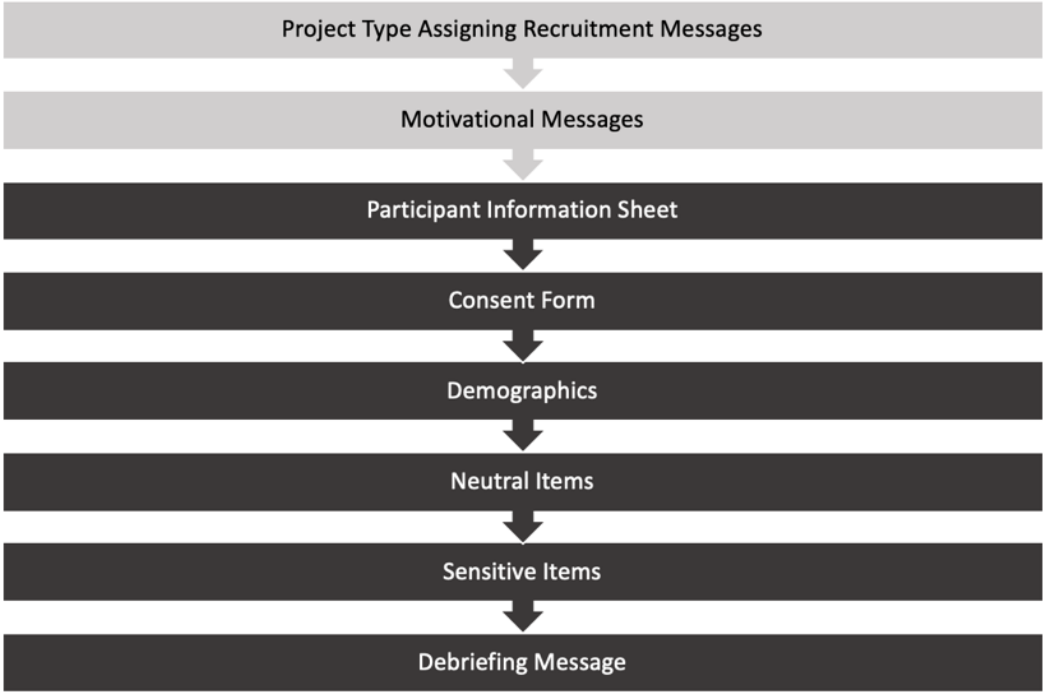


Figure 8 Elements of the Sleep Patterns survey

In Study 3, the independent variables were Project Type and Motivation. Project Type had two levels (Citizen science, and Online quiz) and it was operationalised by recruiting participants with two different twitter messages, one referring to citizen science, and the other referring to an online quiz. Motivation had 3 levels (Learning about science, Learning about self, Control). It was operationalised by presenting the participants with one of the three randomly assigned motivational messages – these were the same messages as the ones used in Study 2.

4.4.3. Results

4.4.3.1. Participant attrition

Recruitment for this study took place in two stages.

4.4.3.1.1. Participant attrition in the citizen science group

Firstly, we recruited participants for the citizen science group. A total of 297 individuals decided to take part in the survey, by agreeing to the main consent question. The attrition of participants throughout the survey is illustrated in Figure 9.

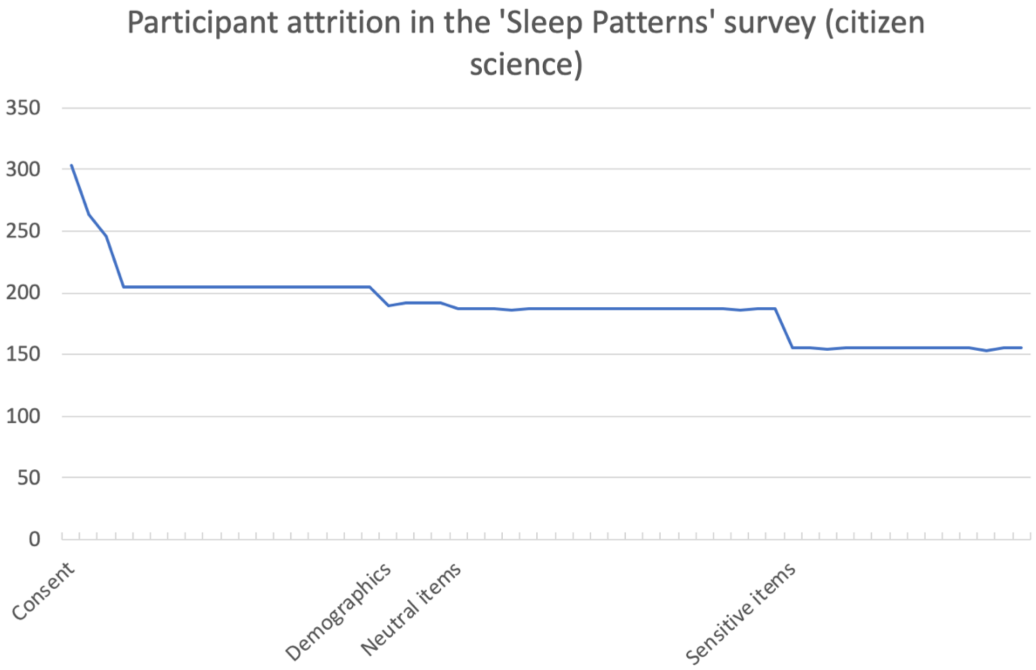


Figure 9 Participant attrition across survey stages in Study 3 (citizen science condition)

Attrition over the 17 secondary consent questions resulted in a sample of n = 190 at the beginning of Demographics items. A sample of 187 answered the first question requesting neutral data, and no attrition occurred over the course of Neutral Items. A decline in participant numbers was then observed at the point of the first question requesting sensitive data: out of 187 participants who answered the last Neutral Item, only 155 answered the first Sensitive Item. This 17.1% 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.

No participants dropped out over the course of Sensitive Items. In total, a sample of n = 155 participants completed the last item of the survey.

Although individuals who followed the link to the survey were randomly assigned to three Motivation groups in equal numbers, due to the pattern of attrition in this sample, the final sample included slightly uneven numbers of participants in the 3 motivational groups: the ‘Learn about science’ group (n = 57), the ‘Learn about self’ group (n = 62), and the ‘Control’ group (n = 36).

4.4.3.1.2. Participant attrition in the online quiz group

Secondly, we recruited participants for the online quiz group. We continued recruitment until we were able to obtain a final sample of n = 155 participants who reached and completed the last survey item, to achieve equal numbers of participants for both levels of the ‘Project Type’ variable.

A total of 450 individuals decided to take part in the survey, by agreeing to the main consent question. The attrition of participants throughout the survey is illustrated in Figure 10.

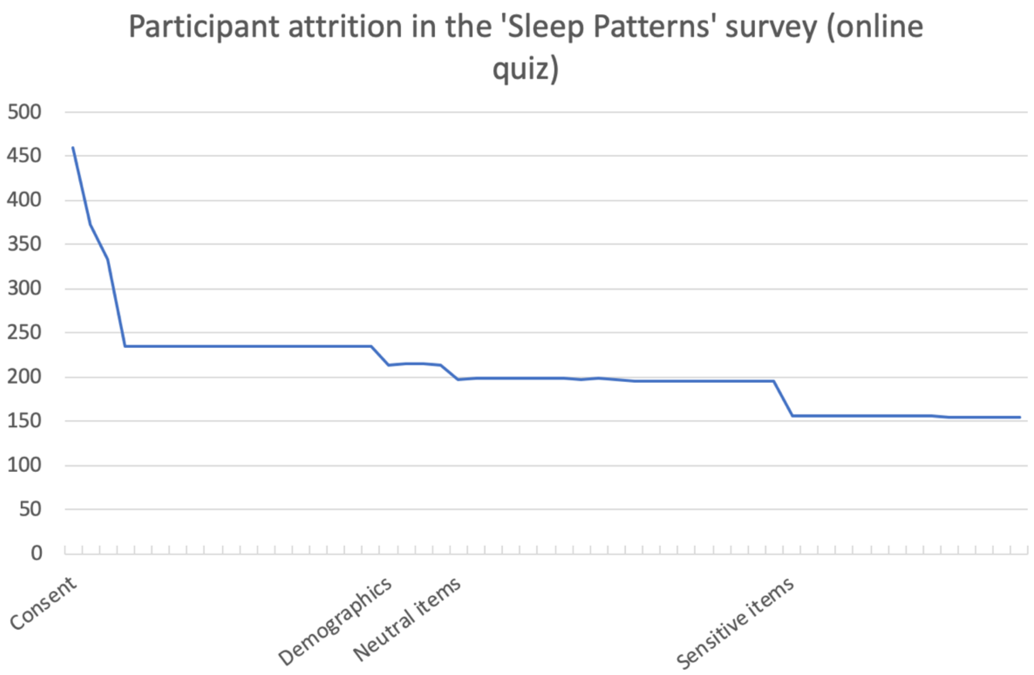


Figure 10 Participant attrition across survey stages in Study 3 (online quiz condition)

Attrition over the 17 secondary consent questions resulted in a sample of n = 213 at the beginning of Demographics items. A sample of 197 answered the first question requesting neutral data, and 1 participant dropped out over the course of Neutral Items. A decline in participant numbers was then observed at the point of the first question requesting sensitive data: out of 196 participants who answered the last Neutral Item, only 156 answered the first Sensitive Item. This 20.4% 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. One participant dropped out over the course of Sensitive Items. In total, a sample of n = 155 participants completed the last item of the survey.

Although individuals who followed the link to the survey were randomly assigned to three Motivation groups in equal numbers, due to the pattern of attrition in this sample, the final sample included slightly uneven numbers of participants in the 3 motivational groups: the ‘Learn about science’ group (n = 56), the ‘Learn about self’ group (n = 51), and the ‘Control’ group (n = 48).

4.4.3.2. Demographics

The total sample consisted of 310 participants.

Data about age, as reported by the participants, were available for n = 308, with ages ranging from 18 to 75 years old (mean = 40, SD = 15).

Data about gender, as reported by participants (available for n = 310) were as follows: n = 177 female, n = 122 male, n = 7 non-binary, n = 1 preferred not to answer, and n = 3 answered in their own words.

Data about Internet usage were available for 310 participants: n = 152 reported using the Internet ‘all the time’, n = 148 reported using the Internet ‘several times per day’, n = 8 reported using the Internet ‘most days’ and n = 2 reported using the Internet ‘several times per week’.

Only a small proportion of participants reported having had previous experience with citizen science (n = 17, ‘once’, and n = 12, more than once’), while the majority reported either not having any citizen science experience (n = 225) or being ‘not sure’ (n = 55); data for one participant were missing.

4.4.3.3. Data disclosure across Neutral Items

All participants answered every of the 19 Neutral Item questions, resulting in 100% neutral data disclosure in this study.

4.4.3.4. Data disclosure across Sensitive Items

Disclosure of sensitive data in this study ranged from 0 to 13 disclosed items (out of 14), with a mean of 4.93 items (SD = 1.96), equivalent to 35.2% disclosure of sensitive data. There was a small positive correlation of data disclosure with the age of participants (r = .215, n = 308, .000), suggesting that older citizen scientists are more likely to share a greater volume of sensitive data. As the counts were uneven in this study (19 Neutral Items and 14 Sensitive Items), they were transformed into percentages before a comparison between the two was conducted. The result of a paired t-test (t = 81.644, df = 309, p = .000, 2-tailed) suggested that the disclosure of Neutral Items was significantly higher than the disclosure of Sensitive Items.

4.4.3.5. Hypothesis 1: Project Type

This study involved two recruitment modes: citizen science and an online quiz. In the citizen science group (n = 155), participants disclosed a mean of 5.2 Sensitive Items (SD = 2.24). In the online quiz group (n = 155), participants disclosed a mean of 4.66 Sensitive Items (SD = 1.59). A one-way between-subjects Analysis of Variance showed that the impact of Project Type on the disclosure of sensitive data was significant ($F(1,308) = 6.045$, $p = .014$, $\eta = .019$), suggesting that individuals recruited for a citizen science project disclose a larger volume of sensitive data than individuals recruited for an online quiz. Hypothesis 1 was accepted.

Moreover, the variable of Project Type interacted with the variable of Citizen Science Experience ($F(3,301) = 4.686$, $p = .003$, $\eta 2 = .045$). Citizen Science Experience had a significant impact on the volume of sensitive data disclosure only among participants in the citizen science group ($F(3,151) = 4.266$, $p = .006$, $\eta = .078$).

4.4.3.6. Hypothesis 2: Impact of Motivation across Project Type conditions

This study aimed to clarify whether the relationship between motivational messages and disclosure of sensitive data varies between the two types of projects: citizen science and an online quiz. There was a significant interaction effect between the variables of Motivation and Project Type ($F(2,304) = 3.568$, $p = .029$, $\eta 2 = .023$), suggesting that motivational messages may differently impact people’s willingness to share sensitive data, depending on how these participants were recruited.

In the Citizen Science condition (n = 155), the mean disclosure of sensitive data across the motivational groups was as follows: m = 5.10 items (SD = 1.99) in the 'Learn about science' group (n = 57), m = 5.61 items (SD = 2.68) in the 'Learn about self' group (n = 62), and m = 4.64 items (SD = 1.59) in the 'Control' group (n = 36). A One-Way Independent Analysis of Variance was not significant (F(2,152) = 2.276, p = .106, η = .029) suggesting a lack of the impact of the Motivation variable on data disclosure.

In the Online Quiz condition (n = 155), the mean disclosure of sensitive data across the motivational groups was as follows: m = 5.09 items (SD = 1.93) in the 'Learn about science' group (n = 56), m = 4.33 items (SD = 1.44) in the 'Learn about self' group (n = 51), and m = 4.50 items (SD = 1.17) in the 'Control' group (n = 48). A One-Way Independent Analysis of Variance (F(2,152) = 3.472, p = .034, η = .044) suggested that the variable of Motivation had a significant impact on the level of data disclosure. A post hoc Tukey test showed that participants in the 'Learn about science' group shared a significantly larger volume of sensitive data than participants in the 'Learn about self' group (p = .036). These results suggest that, in the online quiz context, encouraging people to learn about science is more effective at eliciting disclosure of data, than encouraging people to learn about themselves.

Moreover, we examined the disclosure of sensitive data across the three motivational groups for the whole sample of participants (n = 330). The mean disclosure for these three groups was as follows: m = 5.09 items (SD = 1.95) in the 'Learn about science' group (n = 113), m = 5.03 items (SD = 2.29) in the 'Learn about self' group (n = 113), and m = 4.56 items (SD = 1.36) in the 'Control' group (n = 84). A One-Way Independent Analysis of Variance was not significant (F(2,307) = 2.099, p = .124, η = .013), providing further evidence that in this study the motivational messages accounted for the pattern of disclosure only in the online quiz condition. Hypothesis 2 was accepted.

4.4.4. Discussion

4.4.4.1. Hypothesis 1: Project Type

This study involved two modes of recruitment: one that asked people to take part in an online quiz and find out their sleep score, and one that asked them to participate in a citizen science survey. Although both groups filled the same questionnaire, which made a mention of citizen science (the last Demographics question asks: '*Have you ever taken part in a Citizen Science project before?*') – in line with our hypothesis, the citizen science group shared a significantly larger volume of sensitive data than the online quiz group. Moreover, while we ensured that the final participant numbers in both groups were even (n = 155 in each group), the participant attrition was greater in the online quiz group (a primary sample of n = 450 signed up for the study) than in the citizen science group (a primary sample of n = 297 signed up for the study). These findings suggest that when people decide to fill an online survey because they wish to take part in citizen science, they are more likely to share data within the project and less likely to drop out.

Two types of implications follow from these findings. Firstly, it appears that emphasising the opportunity for self-discovery at the recruitment stage is not a useful strategy to recruit participants into citizen science surveys that involve the disclosure of sensitive data. It is possible that people associate citizen science projects with trustworthy public institutions and therefore feel more comfortable sharing information when the concept of citizen science is emphasised. It is also possible that individuals focused on self-discovery will prioritise their own goals and protect personal data when possible (privacy protection behaviours have been observed in people's interaction with private companies, (Malheiros, Preibusch, & Sasse, 2013)). The exact reasons for the differences in disclosure behaviour between citizen science and online quiz groups should be explored in future research.

Secondly, these findings have ethical implications: it appears that a mention of the citizen science ‘brand’ at the stage of recruitment can lead people to share a larger volume of sensitive data than they would if they were approached to fill a quiz or survey that is not associated with the concept of citizen science. The breadth and variety of online citizen science projects could make it possible for bad actors to obtain sensitive data from people simply by labelling a quiz or other disclosure request a ‘citizen science project’. This puts a responsibility on citizen science coordinators – whose institutions stand to benefit from the volunteer contributions of citizen scientists – to help ensure that the brand of citizen science is not misused. This could be done by developing a system of certification for citizen science projects, or by helping the wider public understand how to check the authenticity of a project.

Finally, only in the citizen science condition, previous experience with citizen science had a significant impact on the volume of disclosed sensitive data (even though there were more individuals with previous citizen science experience in the online quiz condition), further confirming that people approach data disclosure differently when they believe that they are taking part in a citizen science project.

4.4.4.2. Hypothesis 2: Impact of Motivation across Project Type conditions

As predicted, we found that the relationship between the motivational message presented to participants and the volume of sensitive data they disclose is mediated by the type of project that these participants signed up for in the first place. In a citizen science context, there were no significant differences between the motivational groups, suggesting that the type of motivational messages seen by the participant did not influence their data disclosure behaviour. In an online quiz context, encouraging people to learn about science was more effective at eliciting disclosure of sensitive data, than encouraging people to learn about themselves. These results demonstrate that when communicating to potential participants about the benefits of taking part in a project, it is not only the content of the message that can influence their behaviour, but also the stage at which that message is communicated.

The results in the citizen science condition match the results of Study 2. It therefore appears that the participants in Study 2 viewed the survey as a citizen science project, despite the fact that the most popular tweet in that study did not mention citizen science. It follows that the low mean data disclosure in Study 2 could be linked to the inclusion of the motivational question. Whether asking citizen scientists about their motivational preferences can lead to lowered disclosure of data should be further examined.

In the online quiz condition, the participants disclosed a significantly smaller volume of sensitive data when they were presented with a message that encouraged them to discover their sleep score (than participants encouraged to learn about psychology). While the Participant Information Sheet specified that we were looking at the connections between stress and sleep (and therefore justified the presence of both types of questions, sensitive and neutral), individuals prompted to discover a sleep score may have been keen to find it out fast and paid less attention to reading the Participant Information Sheet – and could have been less accepting of sensitive data requests as a result. It is also plausible that these individuals decided to solely part with the data they perceived as necessary to calculate a sleep score (Neutral Items) and not the Sensitive Items described in the survey as questions about stress.

4.4.5. Limitations

While this study provides further insights into the relationship between motivational messages and data disclosure, these findings cannot be applied to all citizen science settings. Many citizen science projects recruit participants from existing communities, built around an interest in a topic area or an interest in citizen science. Study 3 recruited citizen scientists from amongst Twitter users willing to take a citizen science survey about sleep habits – while these individuals have some things in common (for example interest in sleep research, or even interest in citizen science in general), they were not members of any existing community. As a result, they may not have had the opportunity to experience a feeling of belonging or meaningful social interaction with other citizen scientists.

Similar to Studies 1 and 2, Study 3 provides insights that can be applied to participants recruited for one-off contributions, for example via platforms such as LabintheWild (https://labinthewild.org) that host many one-off contribution projects on one website. Nevertheless, more research needs to be conducted to clarify the impact that community membership may have on the relationship between motivation and data disclosure in citizen science.

4.4.6. Conclusion

This study demonstrated that people tend to share a larger volume of sensitive data when they believe that they are taking part in a citizen science project, compared to people who believe that they are filling an online quiz focused on allowing them to discover a personal score. Moreover, the effectiveness of post-recruitment motivational messages in this study was influenced by the type of project the participants signed up for. While in the citizen science scenario, there were no significant differences between the motivational groups, in the online quiz scenario, participants encouraged to learn about science shared a significantly larger volume of data than those encouraged to learn about themselves. It appears that focusing the communication with participants, both pre- and post-recruitment, on self-discovery rather than on learning about science, may negatively impact data disclosure in online surveys.

Section 4.5. discusses the collective findings from Studies 1, 2, and 3.

4.5. General discussion: Studies 1, 2, and 3

Studies 1, 2, and 3 have consistently demonstrated that citizen scientists are far less keen to share information when faced with sensitive data requests than when asked to disclose non-sensitive (neutral) data. Moreover, the information about the project presented to citizen scientists at different stages of participation can influence the volume of data that they disclose. In Study 1, we found that a message about learning opportunities was associated with a larger volume of disclosure of sensitive data, than messages emphasising contribution or social factors.

In Study 2, neither the motivational messages emphasising learning nor the participants’ motivational preferences could account for the pattern of data disclosure among the citizen scientists.

Study 3 demonstrated that when people join a citizen science project, they are prepared to share a larger volume of data than when they join an online quiz. Moreover, comparison of the volume of data in the citizen science condition in Study 3 as well as in Study 1 to the volume of disclosed data in Study 2, the only one which included a motivational question, suggests that asking people about their motivational preferences can make them less likely to share data within projects.

While Studies 1-3 provided insights into the usefulness of motivational messages for data disclosure in citizen science, the applicability of these findings to other contexts is subject to some limitations. These studies were conducted in the context of a survey that looked at people’s sleeping habits. People’s interest in sleep habits, and paid Twitter advertisements, were the basis for attracting participants. Importantly, these individuals were not part of any existing online community. While such participants are sometimes recruited for citizen science projects in the area of psychology (an example of that are projects hosted on the LabintheWild platform, https://labinthewild.org), most citizen science projects tap into or create a community of potential participants, who may contribute to citizen science with varying intensity. The communities may be organised around interest in specific subject areas (for example botany) or simply an interest in citizen science in general (as demonstrated by Jennet et al. (Jennett et al., 2014), citizen scientists often ‘dabble’ in different projects before becoming loyal to one of them).

In the case of individuals who took part in Studies 1-3, there was no noticeable interaction between participants and no pre-existing online community. This could make extending the findings from those studies to projects that do tap into existing communities difficult. It is possible that when participants and project coordinators belong to an existing community, the motivations that drive participation could be influenced by this: for example, participation could be driven by the desire to help a member of the community.

Another matter that has not been conclusively resolved in previous studies is the comparison between two motivational devices: motivational messages and a motivational question (which enquires into the participants’ preferences). These two these devices were studied within the same study (Study 2) and so it was not possible to isolate the effect of one. In the next chapter, we aim to address the two shortcomings of Studies 1-3: the sole focus on participants recruited from outside of any existing online communities, and the lack of clarity of the relative impact of prompted and preferred motivation on data disclosure.

The next chapter reports Studies 4 and 5.

Chapter 5

Patterns of data disclosure in citizen science: Motivational messages versus motivational preferences

5.1. Outline

Studies 1, 2, and 3 have demonstrated that the presentation of short motivational messages, as well as information about the project type, but not participants' motivational preferences, can account for the pattern of data disclosure in one-off citizen science surveys. Nevertheless, in contrast to many citizen science projects, Studies 1-3 were conducted with participants recruited online who were not part of any community. Although there exist studies online, especially in the area of psychology, that recruit from outside of existing communities, citizen science projects frequently tap into or create communities of individuals who have common interests and interact with each other. It is important to examine the impact that community membership may have on the motivations that drive citizen scientists to disclose or withhold personal data.

Another issue where we still do not have clarity is the relative impact of two motivational devices – motivational messages and a motivational question – on the disclosure of sensitive data. It is important to conduct a clear-cut comparison of these two devices in one population of participants to be able to determine whether greater disclosure of sensitive data can be achieved by directing participants' focus towards a specific motivation to participate, or whether it is more effective to make participants consider their motivational preferences.

This chapter reports the findings from two studies, Study 4 and Study 5, which address the gaps in knowledge outlined above. Firstly, these two studies were conducted with participants recruited from an existing online community, whose members actively interact with one another regularly: the community of pet owners on Twitter. The online surveys created for these studies resembled those presented in the previous chapter. However, instead of questions about sleep habits, we created 14 questions about pet ownership, which played the role of Neutral Items. Secondly, Studies 4 and 5 serve the aim of providing a comparison between two motivational devices – randomly assigned motivational messages, and a motivational question that enquires into people's motivational preferences – with participants recruited from the same online community.

In Study 4, participants were presented with four randomly assigned motivational messages. There were no significant differences in data disclosure between the four motivational groups. In Study 5, participants indicated their motivational preferences, by stating which one of the four listed benefits of citizen science should be prioritised in future projects: we found that those who indicated learning and social interaction as their preferred benefits shared a larger volume of data than those who chose altruism or contribution. Collectively, these findings suggest that, in the context of an existing online community, motivational preferences, but not motivations prompted through messages presented to participants, can account for the pattern of data disclosure.

5.2. Study 4 *Pet Owners*: The impact of motivational messages on data disclosure in a community-based citizen science project

5.2.1. Introduction

Previous research suggests a link between motivation to take part in citizen science and willingness to disclose data in that context. The studies presented so far in this thesis demonstrated that the relationship between motivation and data disclosure can change depending on the way a project is advertised to the participants or on whether participants' motivation is manipulated or enquired about.

Studies 1, 2, 3 were conducted with participants who were not part of any existing online community, but rather a random sample of Twitter users who were interested in taking part in a citizen science survey. It should be noted that many citizen science projects rely on established communities of interest (for example bird watchers or people interested in astronomy) or even simply a developed interest in citizen science (platforms such as Zooniverse or LabintheWild allow people to take part in multiple citizen science projects). It is plausible that, in community-based citizen science settings, decisions about whether to disclose or withhold data could be influenced by community ties; an example of that would be Bowser et al.'s participants who spoke of sharing data to accomplish a common goal (Bowser et al., 2017). Therefore, it is important to examine whether the relationship between data disclosure and motivation, demonstrated in studies 1-3, is robust and extends to citizen scientists recruited from existing online communities.

In Study 4, called *Pet Owners*, participants were presented with one of the four randomly assigned motivational messages (the same motivational messages as in Study 1), however – in contrast to Study 1 – participants were recruited from an existing online community of pet owners. The current study sets out to test the following hypothesis:

H1: The motivational messages presented to the participants will have an impact of the volume of sensitive data disclosed within the survey.

5.2.2. Independent Variable in Study 4

Figure 11 depicts the elements of the *Pet Owners* survey.

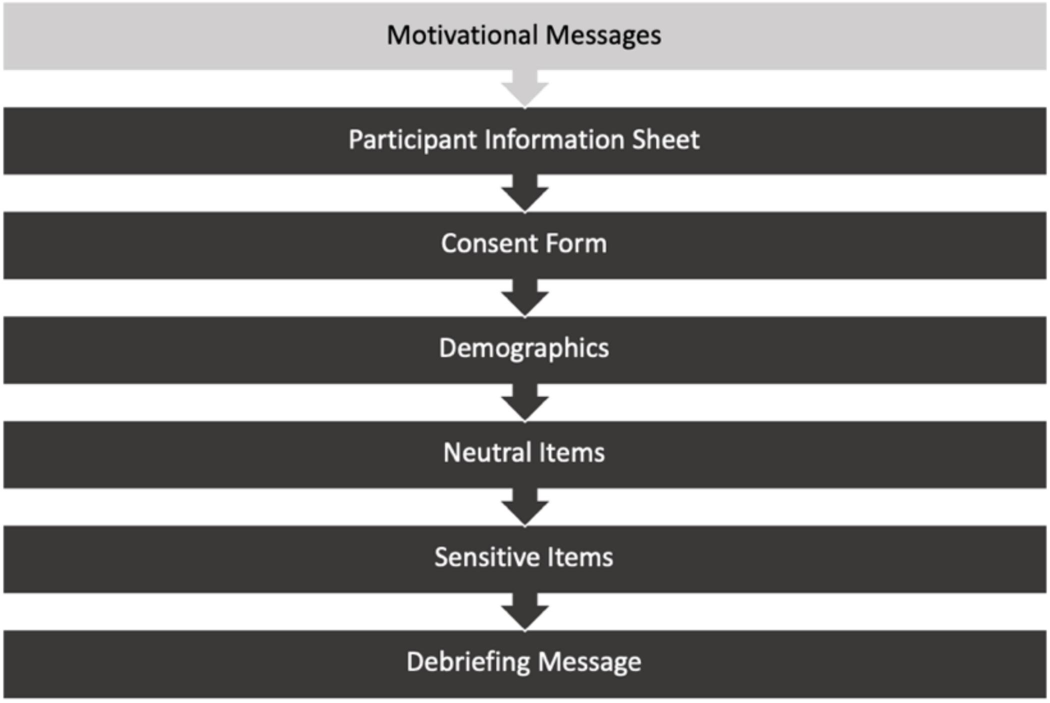


Figure 11 Elements of the Pet Owners survey

The independent variable in Study 4 was Motivation. Motivation had 4 levels (Learning, Altruism, Contribution and Social Proof). It was operationalised by presenting the participants with one of the four randomly assigned motivational messages: '*Extend your knowledge of health psychology by participating in the Pet Owners survey!*' (Learning); '*Join your fellow citizen scientists in establishing connections between stress and pet ownership behaviours!... Many citizen scientists are already participating in the project…*' (Social Proof); '*You can contribute to science by answering questions about sources of stress in your life* *and pet ownership behaviours!*' (Contribution); '*Health psychology needs your help to connect sources of stress to pet ownership behaviours!*' (Altruism).

5.2.3. Results

5.2.3.1. Participant attrition

A total of 503 individuals decided to take part in the survey, by agreeing to the main consent question. The attrition of participants throughout the survey is illustrated in Figure 12.

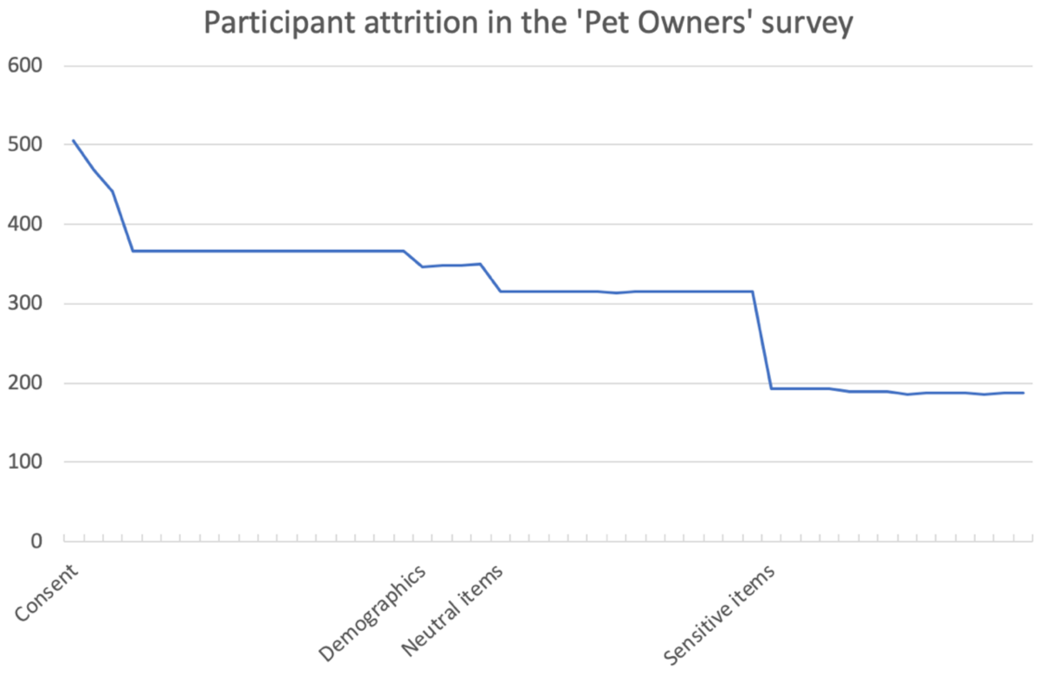


Figure 12 Participant attrition across survey stages in Study 4

Attrition over the 17 secondary consent questions resulted in a sample of n = 347 at the beginning of Demographics items. The last Demographics question was answered by n = 349 (the negative attrition resulting from two participants not answering the first Demographics question). There was a small drop in the number of participants at the point of the first Neutral Item resulting in a sample of 315 individuals. No attrition occurred over the course of Neutral Items.

As in previous studies, a decline in participant numbers was then observed at the point of the first question requesting sensitive data: out of 315 participants who answered the last Neutral Item, only 193 answered the first Sensitive Item. This 38.7% 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.

Additional 5 participants dropped out over the course of Sensitive Items. In total, a sample of n = 188 completed the last item of the survey.

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, the final sample (n = 188) included slightly uneven numbers of participants in the 4 motivational groups: Learning (n = 46), Social Proof (n = 51), Contribution (n = 51), and Altruism (n = 40).

5.2.3.2. Demographics

The final sample (defined as having reached and answered the last question in the survey) consisted of 188 participants.

Data about age, as reported by the participants, were available for the whole sample, with ages ranging from 20 to 78 years old (mean = 45, SD = 12).

Data about gender were as follows: 160 participants reported being female, 24 reported being male, 3 participants reported being non-binary, and 1 participant preferred not to state their gender.

When asked about their use of the Internet, 105 participants stated that they use the Internet ‘all the time’, 77 reported using the Internet ‘several times per day’, while 6 participants stated that they use the Internet ‘most days’.

In response to the question about previous experience of citizen science, 33 participants reported having taken part in a citizen science project in the past (n = 20 ‘once’, and n = 12 ‘more than once’), 130 reported that they had no previous experience with citizen science, whereas 25 stated that they were unsure whether or not they had previous experience of citizen science participation.

5.2.3.3. Data disclosure across Neutral Items

Neutral Item disclosure ranged from 12 to 14 items (mean = 13.78, SD = .51), equivalent to 98.4%.

5.2.3.4. Data disclosure across Sensitive Items

Sensitive Item disclosure ranged from 1 to 13 items (mean = 5.24, SD = 2.13), equivalent to 37.4%. The result of a paired t-test (t = 54.688, df = 187, p = .000, 2-tailed) suggested that the disclosure of Neutral Items was significantly higher than the disclosure of Sensitive Items.

5.2.3.5. Hypothesis 1: Motivation

Participants were divided into 4 motivational groups, based on which one of the 4, randomly assigned, motivational messages were presented to them before the beginning of the survey. The largest volume of sensitive data disclosure was demonstrated by those in the Altruism group (mean = 5.52, SD = 1.93), followed by the Learning group (mean = 5.35, SD = 2.28), the Social group (mean = 5.09, SD = 1.85) and the Contribution group (mean = 5.06, SD = 2.40). The differences between motivational groups were not significant (F(3,184) = .474, p = .701, η = .008). Hypothesis 1 was rejected.

5.2.4. Discussion

5.2.4.1. Data disclosure

Similar to the previous three studies, participants rarely withheld information when responding to Neutral Item questions (with mean disclosure equivalent to 98.4%), a much higher level of disclosure than the mean disclosure of Sensitive Items (equivalent to 37.4%). These results were comparable to the previous three studies, where the disclosure of neutral data was very high (99.9%, 100%, and 100%), while the disclosure of sensitive data was drastically lower (38.4%, 34.5%, and 32.5%, respectively). Collectively, these findings demonstrate that citizen scientists respond differently to requests for neutral data and sensitive data and that this finding is not influenced by whether or not the participants are recruited from an established online community.

5.2.4.2. Hypothesis 1: Motivation

Contrary to our predictions, the variable of Motivation did not have a significant impact on the volume of the sensitive data disclosed by the participants in Study 4. The highest level of sensitive data disclosure was demonstrated by participants who saw the message emphasising the opportunity to help scientists (Altruism). Nevertheless, the difference between the Altruism group and the other three motivational groups was not significant. Because of the nature of the population from which the participants were recruited (an established online community of pet owners), it is plausible that many individuals without an established interest in citizen science would still take part to support the researcher who is also a member of this community. However, as the impact of the variable of Motivation on data disclosure was not significant, it is impossible to state whether focusing citizen scientists' attention on Altruism may be an effective way to support disclosure of sensitive data in existing online communities.

Study 1, *Sleep Mapping*, and the current study, Study 4, employed the same four motivational messages. In Study 1, the variable of Motivation had a significant impact on the disclosure of sensitive data, and the differences between motivational groups stemmed from the higher disclosure in the Learning group. It is likely that the *Sleep Mapping* project attracted a sample of participants with a specific interest in sleep research, and therefore more likely to share sensitive data when their attention was brought to the learning opportunities that may stem from filling the survey. The current study recruited participants from a more generic population, that of Twitter users who own a pet. The relationship between motivation to learn and willingness to share data may be specific to projects with a narrower topic focus.

The population of pet owners on Twitter appear to value community-relevant goals like helping and social interaction (for example, they comment on one another’s posts and demonstrate helpfulness by amplifying others’ messages and organising support when a community member is in need). When recruiting from established online communities, building projects that are grounded in community-relevant values could be a first step towards encouraging data disclosure. A solution worth exploring could be liaising at the stage of study design, perhaps through cooperation with members of the community who do wish to become involved in citizen science and participatory research. Such cooperation could be beneficial for both the researchers and the community, the first benefitting from access to an existing pool of potential participants, the second being able to bring to light research questions they care about. Indeed, it is likely that projects which require disclosure of sensitive data will need to work closely with online communities to build trust and cooperation.

5.2.5. Limitations

As with all studies presented in this thesis, the current study employed a questionnaire that was originally developed for research in a credit card application context. It is possible that the high concentration of questions requesting sensitive data was unexpected for some of the participants in this study and may have caused them to drop out. When participants are recruited due to reasons unrelated to science (e.g. being a pet owner as opposed to having an interest in sleep research), the sample may be more surprised at a high number of questions requesting sensitive data. Future investigation into data disclosure in citizen science settings would benefit from research that maps out and systemises the sensitive data requests, and patterns of responses, across existing and past projects. While likely to entail significant challenges (such as obtaining the datasets), systematic analysis of information disclosure in existing projects could inform the design of studies to examine the disclosure behaviour of citizen scientists.

5.2.6. Conclusion

In Study 4, participants were randomly assigned to 4 motivational groups. In contrast to Study 1 (which utilised the same motivational device) the level of disclosure of sensitive data in Study 4 did not significantly differ across the motivational groups. While the Altruism motivation was the least effective at retaining participants within the study, it was associated with the highest level of sensitive data disclosure, suggesting that motivation to complete a survey and motivation to share sensitive data within that survey, do not always align in citizen science.

It is important to clarify the distinction between the impact of motivational messages and motivational preferences on data disclosure. Future studies should explore the relationship between preferred and randomly assigned motivations to establish whether citizen science projects should be designed to fulfil participants’ preferences or whether more attention could be placed on drawing participants’ attention to already existing benefits of taking part. To bring clarity to this matter, Study 5 investigates the impact of motivational Preference (operationalised by asking participants to choose one of the four benefits of citizen science that should be prioritised in future projects) among participants drawn from the same online community that served as a recruitment pool in Study 4. Together, these two studies will allow a comparison between the relative impact of motivational messages and motivational preferences, on the volume of sensitive data disclosed by citizen scientists.

5.3. Study 5 *Pet Experience*: The impact of motivational preferences on data disclosure in a community-based citizen science project

5.3.1. Introduction

Study 4 has demonstrated that, in a sample of participants recruited from an active online community, motivation prompted by short messages presented at the start of the survey could not account for the pattern of disclosure of sensitive data. It is possible that these individuals took part in the study to support a member of the community and therefore were not receptive to randomly assigned motivational messages. It would be worthwhile, however, to examine whether asking about their motivational preferences might more effectively bring their attention to the different benefits of taking part in the survey – and thus have an impact on the volume of sensitive data that they decide to disclose. In a study of experienced citizen scientists, researchers found that people's own motivations and priorities account for their willingness to disclose data (Bowser et al., 2017).

Study 5, named *Pet Experience*, looks at the disclosure of data in a sample of pet owners. Similar to Study 2, participants were asked to answer a motivational question, choosing one of the four benefits of citizen science that they believe should be prioritised in future citizen science projects. In contrast to Study 2, however, participants were not presented with any motivational messages. We set out to clarify the extent to which people's preferences about the benefits of citizen science may account for the volume of sensitive data that they are willing to disclose.

We anticipate that the variable of Preference will have a significant impact on the disclosure of sensitive data. The current study sets out to test the following hypothesis:

H1: Participants' motivational preferences will have an impact of the volume of sensitive data disclosed within the survey.

5.3.2. Independent Variable in Study 5

Figure 13 depicts the elements of the Pet Experience survey.

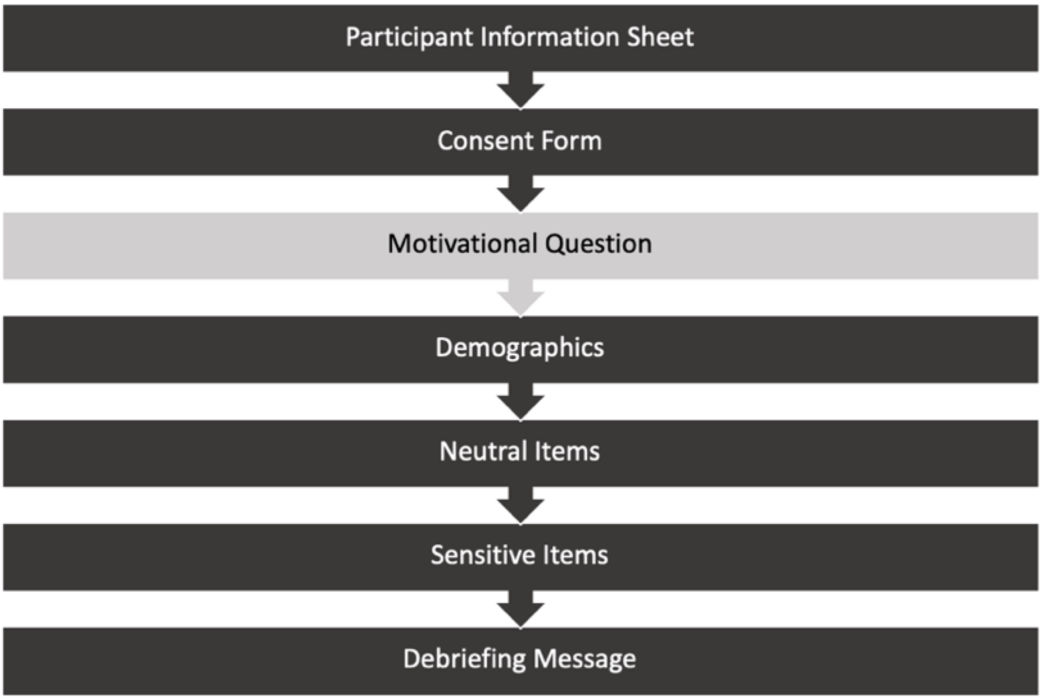


Figure 13 Elements of the Pet Experience survey

In Study 5, the independent variable was Preference. Preference had 4 levels (Learning, Altruism, Contribution and Social interaction). It was operationalised by asking participants to answer a motivational question that stated: '*Before filling the survey, please answer this question to help us design better citizen science projects: Which benefit of taking part in citizen science is most important to you?*'. Participants were given four options from which to choose one: 1) Learning about science, 2) Social interaction, 3) Making a contribution to a project, 4) Helping scientists.

5.3.3. Results

5.3.3.1. Participant Attrition

A total of 364 individuals decided to take part in the survey, by agreeing to the main consent question. The attrition of participants throughout the survey is illustrated in Figure 14.

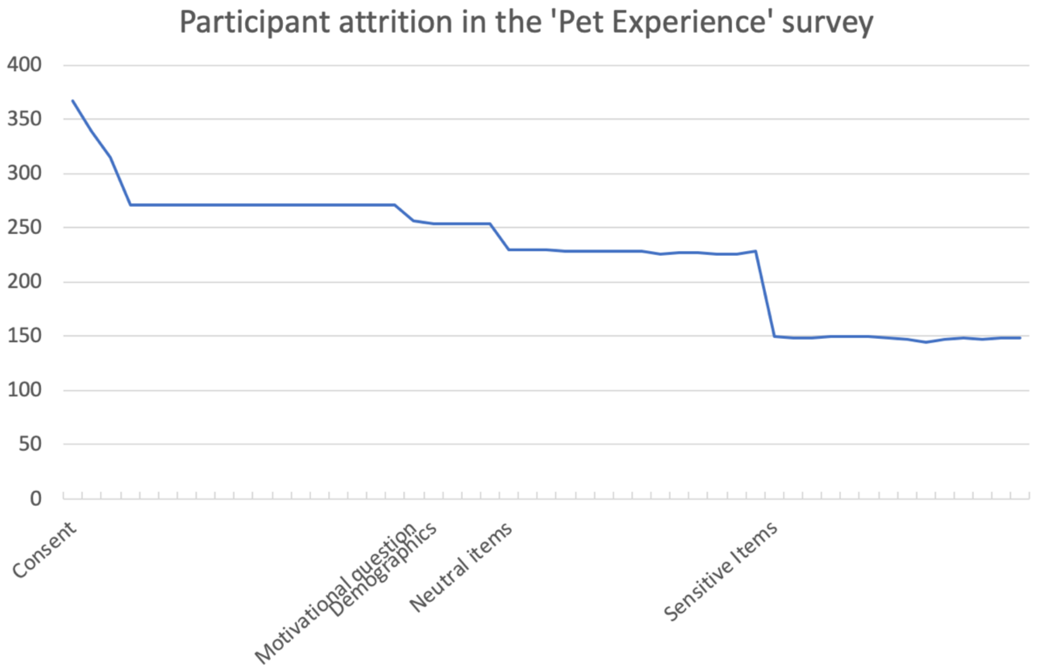


Figure 14 Participant attrition across survey stages in Study 5

Over the course of consent questions, 110 participants dropped out, resulting in a sample of 256 participants who answered the motivational question. There was a smaller point of attrition between the last Demographics item (n = 254) and the first Neutral Item (n = 230). Only 2 further participants dropped out over the course of Neutral Items. A decline in participant numbers was then observed at the point of the first question requesting sensitive data: out of 228 participants who answered the last Neutral Item, only 150 answered the first Sensitive Item. This 34.2% 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.

Two participants dropped out over the course of Sensitive Items. In total, a sample of n = 148 participants completed the last item of the survey.

5.3.3.2. Demographics

The final sample (defined as having reached and answered the last question in the survey) consisted of 148 participants.

Data about age, as reported by the participants, were available for n = 143, with ages ranging from 18 to 78 years old (mean = 40, SD = 14).

Data about gender, as reported by the participants (available for n = 147) were as follows: n = 114 female, n = 32 male, n = 1 preferred not to disclose gender information.

Concerning Internet usage, 97 participants stated that they used the Internet ‘all the time’, 49 used the Internet ‘several times per day’, while 2 participants stated that they used the Internet ‘most days’.

In regards to previous experience with citizen science, a sample of n = 10 reported having participated ‘once’, and n = 14 reported having participated ‘more than once’, while the majority of participants in this study did not report having any previous experience with citizen science (n = 106 answered ‘no’, and n = 18 said they were ‘not sure’).

5.3.3.3. Data disclosure across Neutral Items

Neutral Item disclosure was very high and ranged from 8 to 14 items (mean = 13.66, SD = .82), equivalent to 97.6%.

5.3.3.4. Data disclosure across Sensitive Items

Sensitive Item disclosure ranged from 1 to 14 items (mean = 4.82, SD = 1.67), equivalent to 34.4%. The result of a paired t-test (t = 60.641, df = 147, p = .000, 2-tailed) suggested that the disclosure of Neutral Items was significantly higher than the disclosure of Sensitive Items.

There was a positive correlation between age and Sensitive Item disclosure (r = .262, n = 143, p = .002), suggesting that older citizen scientists are more likely to share a greater volume of sensitive data.

5.3.3.5. Hypothesis 1: Preference

In this study, participants self-assigned to motivational Preference groups by choosing one of the 4 benefits of citizen science, which they thought should be prioritised in the future (Learning, Social interaction, Contribution, or Helping/Altruism). Based on the choices made by the participants, and the attrition following the motivational question, the four motivational groups were uneven: Learning (n = 13), Social (n = 7), Contribution (n = 77), and Altruism (n = 51).

The highest Sensitive Item data disclosure was demonstrated by the Social group participants (mean = 6.57, SD = 3.41), followed by the Learning group (mean = 5.31, SD = 1.75), Contribution (mean = 4.73, SD = 1.55) and Altruism (mean = 4.61, SD = 1.37). A significant result of a between-groups Analysis of Variance (F(3,144) = 3.441, p = .019, η = .067) indicated that Preference had a significant impact on Sensitive Item disclosure. Hypothesis 1 was accepted.

A post hoc Tukey test showed that the Sensitive Item disclosure in the Social group was significantly different from the Sensitive Item disclosure in the Contribution group (*p* = .025), and in the Altruism group (*p* = .017). There was no significant difference on the Sensitive Item disclosure between the Social and Learning, Learning and Contribution, or Contribution and Altruism groups. These results suggest that it was the Social group that was primarily responsible for the impact of the Preference group on sensitive data disclosure.

5.3.3.6. Post-hoc analysis: Egoic versus selfless Preference

Furthermore, although the Learning group disclosure did not differ significantly from the Altruism or Contribution groups, neither did it differ from the (highest-disclosure) Social group, suggesting that there may be an overall difference between motivations that emphasise what one can *gain* from participating in citizen science (learning and social motivations) and motivations that emphasise what one can *give* (altruism and contribution). To explore this possibility, we computed two overarching groups, ‘Egoic Preference’ (composed of participants who prioritised learning and social interaction) and ‘Selfless Preference’ (composed of participants who prioritised altruism and contribution). The Egoic Preference group (n = 20) was found to have higher levels of Sensitive Item disclosure (mean = 5.75, SD = 2.45), than the Selfless Preference group (n = 128) (mean = 4.68, SD = 1.48), and this difference was found to be significant (F(1,146) = 7.387, p = .007, η = .048).

5.3.4. Discussion

5.3.4.1. Data Disclosure

We found that citizen scientists were more likely to share neutral data than sensitive data about themselves or others. This is in line with the research conducted with credit card applicants (Malheiros, Brostoff, et al., 2013), as well with Studies 1, 2, 3, and 4 presented in this thesis.

Study 5 has further supported these findings, showing that citizen scientists do pay attention to the questions they are being asked, and oftentimes are not comfortable with sharing information about third parties. Future citizen science projects, especially those that focus on gathering data about citizen scientists (such as citizen psych-science surveys) are likely to face challenges when requesting sensitive information, even when participant recruitment takes place within an existing online community, in which the sharing of information takes place frequently.

Similar to Studies 2 and 3, older participants disclosed a larger volume of sensitive data, suggesting that older people may be more vulnerable to data disclosure requests online.

Finally, it is interesting to note that the level of sensitive data disclosure in the current study was very similar to that in Study 2, which also employed the motivational question – and mean disclosure in both of those studies was noticeably lower than the mean disclosure in Studies 1, 3 and 4 which did not employ the motivational question. One way to explain the results outlined above could be to argue that, when seeking to increase disclosure of sensitive data, asking citizen scientists to consider their motivational preferences is a less effective motivational technique, than presenting participants with randomly assigned messaged. It is even possible that asking people to consider what motivates them the most could make them reflect more sceptically on the value (to them) of the survey they are filling, and therefore share a smaller volume of data.

5.3.4.2. Hypothesis 1: Preference

In this study, individuals who indicated social motivation as most important disclosed the largest volume of sensitive data. Considering the small size of this group (n = 7), it is difficult to state to what extent this finding can be extrapolated from to other citizen science projects. In line with our predictions, citizen scientists who prioritised the rewards of participation that they can gain (such as learning or social interaction, n = 20) disclosed higher mean volumes of sensitive data than those who prioritised the giving aspects of citizen science (altruism or contribution, n = 128). These findings are in line with previous research, which linked primary engagement with a citizen science study to reasons related to egoism such as personal interest in citizen science and curiosity (Rotman et al., 2012). It appears, therefore, that encouraging data sharing among dabblers, or in the context of projects that only require a one-off contribution, should focus on supporting egoic motivational goals.

The findings described above suggest that, when making decisions about parting with their data, citizen scientists need tangible rewards that they can then weigh against the potential risks of participation. It follows then that those who design, and co-ordinate citizen science projects need to ensure that the benefits offered in exchange for participation are attractive and well-matched to citizen scientists' preferences.

5.3.4.3. Post-hoc comparisons: Egoic and Selfless Preference

We computed two overarching variables: Egoic Preference (which included participants who saw the messages focused on Learning or Social Interaction) and Selfless Preference (which included participants who saw the messages focused on Altruism or Contribution). We found that participants prompted by messages that emphasised 'egoic' preferences disclosed a significantly larger volume of sensitive data than those prompted by messages about 'selfless' preferences.

These findings are in line with the existing literature. A study conducted by Rotman et al. found that primary engagement with a citizen science project is driven by egoic motivations such as personal interest in citizen science and curiosity (Rotman et al., 2012).

Furthermore, this study recruited participants for a one-off contribution, rather than asking them to engage in repeated data collection or analysis tasks that are additive and promote sustained contribution over longer periods. Therefore, by design, our participants were 'dabblers', taking part only once. There is evidence to suggest that motivation to contribute to science is stronger among participants who have participated in a project for a longer period (Land-Zandstra, van Beusekom, Koppeschaar, & van den Broek, 2016). Because of this, participants who care about contributing, might not have found this one-off contribution to *Pet Experience* as fulfilling as they would, were they a part of a long-term project.

Furthermore, although filling an online survey does not typically include features that would promote social interaction with other participants, pet owners on Twitter form an established online community. Within this community, they show off photos of their pets, and interact with one another. Therefore, it is possible that due to recruitment of participants from an active online community, those who care about social interaction were motivated to disclose a larger volume of data (for example, the mention of social interaction could remind them of the satisfaction they draw from being part of the community, thus setting the survey in a context where their motivational preferences are met).

Interestingly, in this study, we observed a discrepancy between the most popular (most frequently chosen by the participants) motivations and the motivations that aligned with higher levels of data disclosure. The great majority of citizen scientists in this sample saw the ‘selfless’ benefits of altruism and contribution as more important (n = 128) than the ‘egoic’ learning and social interaction, favoured by a small minority of only 20 participants. The latter motivational preferences, however, were associated with higher levels of data disclosure.

It may be, that the current distinction between intrinsic (stemming from the enjoyment of task) and extrinsic (stemming from the expectation of fulfilling a future goal) motivation, commonly used to describe the reasons that citizen scientists take part in projects (Nov, Arazy, & Anderson, 2011), are insufficient for the study of privacy and data disclosure in citizen science. Future studies should further explore the role of egoic and selfless rewards for eliciting data disclosure in citizen science contexts.

5.3.5. Limitations

The Preference groups in this study were uneven, with the groups that demonstrated significantly higher disclosure (Social and Learning) making up only 13.5% of the total sample. This is an unavoidable issue when studying participants' motivational preferences and therefore allowing participants to self-assign to groups.

Furthermore, it should be noted that, when asking participants about their motivational preferences, the Social answer option emphasised social interaction. This was not an exact match for the Social Proof message in Study 4, which instead informed people about the many individuals who are already taking part in the project (relying on the phenomenon called 'social proof', whereby people are more likely to engage in an action that they know others engage in (Cialdini, 2001). The phenomenon of social proof, however, could appear unnaturally phrased as a benefit of citizen science, and as a result we decided to focus on the benefit of social interaction instead.

5.3.6. Conclusion

This study has confirmed the existence of a link between the benefits of citizen science that participants value the most, and their willingness to disclose sensitive data within a citizen science project. Participants for whom gaining rewards such as learning or social interaction were more important than the more selfless benefits of altruism or contribution, disclosed a larger volume of sensitive data. Factors associated with lower data disclosure were younger age and lack of familiarity with the concept of citizen science. Section 5.4. discusses the collective findings from Studies 4 and 5.

5.4. General Discussion: Studies 4 and 5

This chapter presented two studies that looked at the impact of motivation on the disclosure of sensitive data among citizen scientists recruited from an existing online community. They both used an online survey that consisted of Neutral and Sensitive Items. However, while Study 4 prompted motivation with randomly assigned messages, Study 5 enquired into participants' motivational preferences.

In Study 4, participants were presented with a randomly assigned message, that emphasised one of the four motivations (Learning, Altruism, Contribution, and Social Proof). We found that there were no significant between-group differences in the disclosure of sensitive data – suggesting that in this sample, motivation prompted by randomly assigned messages could not account for the pattern of data disclosure. It is possible that, when participants are recruited from an online community, they may choose to take part simply due to a willingness to help a member of this community (in this case, the research student) and are not much concerned with other benefits of participation or lack thereof. Furthermore, we discovered that while participants prompted by the message emphasising Altruism were the most likely to drop out, those who remained within the survey shared the largest volume of data in comparison to the other three groups. It, therefore, appears that, in the context of one-off citizen science contributions, the motivation to take part and the motivation to disclose sensitive data do not always align – and may need to be studied as two separate concepts.

In Study 5, participants were asked to indicate their motivational preferences, by choosing one of the four benefits of citizen science (Learning, Altruism, Contribution, and Social interaction) that should be prioritised in future citizen science projects. Participants' preferences accounted for the pattern of disclosure: those individuals who chose the benefits associated with what they can gain from participation (learning and social interaction) shared a significantly larger volume of data than those who chose the selfless benefits of altruism and contribution. This is consistent with previous research that showed egoic motivations to be the primary driver at the early stages of citizen science participation (Rotman et al., 2012). Notably, the two motivational preferences associated with a larger volume of disclosure were also the least popular ones, demonstrating that the benefits of citizen science participation that appeal to the widest range of individuals may not necessarily be the ones that most effectively increase disclosure of sensitive data. Finally, the overall data disclosure in Study 5 was lower than in Study 4, suggesting that asking citizen scientists to consider their preferences may cause some participants to withhold sensitive data more frequently, perhaps due to an unfavourable comparison between their own prioritises and the benefits that a particular project offers.

Collectively, Studies 4 and 5 show evidence that, in a sample of participants recruited from an existing online community, participants' preferences can better account for the pattern of data disclosure across different motivations, than randomly assigned motivational messages – however, it is the latter motivational device that results in higher overall disclosure of sensitive data. Moreover, the motivation to take part and the motivation to disclose data are separate concepts in citizen science and, when designing projects, the two may need to be balanced to achieve sufficient levels of both participation and data disclosure.

It is important to consider the possibility that the pattern of data disclosure in Studies 1-3 differs from that in Studies 4 and 5 not because the latter studies drew upon an existing online community, but because in the more general context of a survey about pets, the sensitive questions could be perceived as less acceptable than in the more science-specific context of a survey about sleep habits. This is, however, unlikely considering that the mean volume of disclosed sensitive data in Study 2 (which employed the motivational question in a sleep habits context) was almost identical to that in Study 5 (which employed the motivational question in a pet owner context): 4.84 and 4.82, respectively.

Notably, in the question validation study, reported in Chapter 3, both in the sleep habits condition and in the pet owner condition, all sensitive data requests were accurately perceived by the participants as asking for sensitive data. This, therefore, suggests that it is unlikely that the participants in the pet owner sample perceived the sensitive questions as more invasive than did the participants in the sleep habits sample – instead, it appears that the differences between the data disclosure in the sleep habits sample and the pet owner sample are more pronounced when it comes to the patterns of data disclosure across different motivational groups.

Chapter 6

General discussion

6.1. Introduction

This thesis reports on five studies that investigated the relationship between motivation to take part in citizen science, and the volume of sensitive data shared by participants. In this chapter, we provide an answer to the main research question and discuss how the findings from this research can be put into practice. We also outline the ethical considerations that are important to acknowledge when researching this area.

6.2. The research aim

This thesis aimed to answer the following question:

‘*How does communication about the rewards of citizen science participation affect the willingness of citizen scientists to share personal data?’*

Through conducting five studies, we have demonstrated that communication about the benefits of taking part in a citizen science survey can have a significant impact on the volume of sensitive data disclosed by the participants. Participants recruited from outside of existing communities share a larger volume of data when presented with motivational messages that emphasise learning opportunities, than when encouraged through messages that emphasise contribution or social aspects of participation. However, it appears that messages emphasising the motivation to learn (whether about science or oneself) do not *increase* the volume of data disclosed by citizen scientists when compared to a neutral control message. Moreover, merely being aware that one is taking part in a citizen science survey leads to a larger volume of disclosed data, when compared to taking part in an online quiz survey that does not mention the concept of citizen science – despite both surveys being otherwise identical. In the case of citizen scientists recruited from an established online community, it appears that drawing participants' attention to their motivational preferences can facilitate the disclosure of sensitive data, but only for those participants who favour learning and social interaction.

Overall, for participants recruited from both within and outside of existing communities, presentation of motivational messages results in a larger volume of disclosed data when compared to enquiring about participants' motivational preferences. It appears that directing citizen scientists' attention to their motivational preferences can only cause increased disclosure for a small select group of participants while lowering the volume of data disclosed by others.

Finally, we have demonstrated that the types of communication that can encourage the largest volume of participants to complete a survey will not necessarily result in the largest volume of disclosed data. In other words, we provide evidence that the motivation to take part in citizen science and the motivation to disclose data within citizen science projects are two different types of motivation, and therefore must be treated as distinct both in research studies, as well as in practice.

In sum, we have addressed the research aim by demonstrating that communication about the benefits of taking part in citizen science can impact data disclosure through the use of motivational messages, enquiring about participants' motivational preferences, and emphasising the concept of citizen science at the stage of recruitment. In the following section, we discuss the findings from studies presented in this thesis and contextualise them within the existing literature.

6.3. Discussion of findings

6.3.1. Do dabblers approach disclosure differently than experienced citizen scientists?

Before we conducted the studies reported in this thesis, there had been only one investigation of data disclosure with citizen scientist participants. Bowser and colleagues (Bowser et al., 2017) interviewed project coordinators and experienced citizen scientists and discovered that the motivations which drove people to take part in citizen science overshadowed their privacy concerns. In Studies 1-5, exploring one-off participation in citizen science, we found that the motivation to take part and the motivation to share data do not always align. For example, in Study 1, the Learning motivation was the most effective at eliciting data disclosure, and at the same time the second least effective at retaining participants within the study. This situation was mirrored in Study 4, where participants who saw the Altruism message disclosed the largest volume of data, but it was the smallest of the four groups, demonstrating that individuals who saw the Altruism message were more likely to drop out of the survey. It is possible that some of our participants were motivated to finish the survey, however, they were not aware that not merely completion of the survey, but also the disclosure of data, are necessary for the success of a citizen science project. In contrast, the experienced citizen scientists interviewed by Bowser et al. appeared well aware that sharing data with project coordinators is a building block of scientific investigation (Bowser et al., 2017). Dabblers, having not received citizen science training, may not understand that incomplete data could put the project’s viability at risk. Future research should explore whether providing more information about the aims and requirements of citizen science could help facilitate data disclosure in one-off projects.

It should be noted that the sensitive questions used in our surveys were adapted from a study of disclosure in a credit card environment (Malheiros, Brostoff, et al., 2013). While it was of benefit to utilise a survey that had been validated in a previous study, the personal nature of some questions could have been seen as excessive by some of our participants (more so than if, for example, we included only the sensitive questions that referred overtly to the sleep and pet themes of the surveys). It is, therefore, possible that disclosure of sensitive data is less problematic in the context of citizen science than it would appear based on the results of Studies 1-5. Nevertheless, researchers using citizen science as a tool for data collection in the areas of health and psychology may require a breadth of information that not every participant will understand the relevance of. Future research could further aid the design of citizen science projects by examining participants' perception of the relevance of various sensitive questions in the context of psychological citizen science and the impact that it may have on the volume of disclosed data.

6.3.2. The role of incentives in encouraging data disclosure among dabblers

6.3.2.1. Motivational messages

As discussed in Chapter 2, the design of Studies 1-5 was underpinned by research into the role of incentives in citizen science recruitment. In Study 1, we presented the participants with one of the four different motivational messages (emphasising altruism, contribution, learning or social proof). These messages were adapted from a study of citizen science recruitment (Lee et al., 2018). Lee et al. found that people who saw a message about altruism were significantly less likely to click the link to the project than those who saw one of the other three messages, however, they then provided the largest volume of initial contributions. In Study 1, the Altruism group was the second largest group of participants and also the group with the second highest mean of disclosed Sensitive Items – suggesting that in Study 1 the message focused on altruism was fairly effective at both retaining participants within the study and prompting disclosure of sensitive data.

Notably, the use of motivational messages in our studies demonstrated that willingness to participate and willingness to share data can act as two separate processes among citizen science dabblers. In Study 4, which also employed the same four messages, the Altruism group was the smallest but demonstrated the highest level of data disclosure.

A direct comparison with Lee et al., however, is not possible as they studied all participants who clicked the link to the survey (perhaps due to the additive nature of participation in that project), while we only studied the data of individuals who completed the full survey.

Another study that examined the impact of motivational messages on recruitment looked at slogans used to recruit participants for unpaid scientific experiments (August et al., 2018). The researchers demonstrated that people were more likely to participate when encouraged by slogans that emphasised self-discovery (such as self-learn or compare). In line with this, both in Study 2, as well as in the Citizen Science condition in Study 3, participants who saw a message emphasising learning about oneself, we more likely to complete the survey (these groups experienced less attrition). In the Online Quiz condition, however, the group with the least attrition was the one where participants saw a message emphasising learning about science. This demonstrates that information given at different stages of participation may interact in their effect. Importantly, we found that participants who believed that they were taking part in a citizen science survey shared a significantly larger volume of sensitive data when prompted to learn about science than when prompted to learn about themselves. This could be connected to the primary reasons for why a particular participant responds to a particular message/slogan. August et al. found that people who responded to the self-learn slogan were more motivated by boredom, which, the authors warn, could lead to less attentive participants. Commenting on the results of their study, the researchers caution against recruitment bias, saying that '*The fact that participants with certain motivations were more likely to self-select into studies advertised with specific frames suggests that how a researcher advertises a study can lead to sample biases in terms of participant motivation*' (August et al., 2018).

In the context of projects that require disclosure of personal data, the coordinators will need to strike a balance between their ability to recruit participants into a project as well as ensuring that these participants are willing to share data within the projects. The techniques employed to achieve these goals will also need to be capable of attracting a diverse enough range of individuals, such as to avoid self-selection biases that would limit the validity of the findings. All this could be difficult to accomplish in an area that already struggles with recruitment and retention of participants (Riesch & Potter, 2014). Nevertheless, bringing a greater focus to how the participant can benefit from taking part in the project and communicating about these rewards would be likely to help coordinators address these issues.

6.3.2.2. Motivational preferences

We also asked citizen scientists (in Studies 2 and 5) to indicate which benefits of citizen science should be prioritised in the future. Only in the latter study did people's preferences account for the pattern of data disclosure: participants who prioritised learning or social interaction shared a larger volume of sensitive data than participants who prioritised contribution or altruism. These findings are not surprising, as research suggests that motivation to contribute to science increases with time (Land-Zandstra et al., 2016) and our investigation focused on one-off participation in a survey. In line with this, Rotman et al. found that early engagement in citizen science is driven by egoic motivations like curiosity and personal interest in science (Rotman et al., 2012). It should be noted, however, that the Learning and Social interaction groups, which shared the largest volumes of data, were also the smallest groups. This further demonstrates that motivation to participate and motivation to disclose are two separate processes.

The type of motivational device (message versus question) appears to alter the impact of a particular motivation on the participants – in Study 4, altruism was most effective at eliciting data disclosure but least effective at retaining participants within the study. Conversely, in Study 5, altruism was the second most effective at retaining the participants but was also associated with the lowest disclosure of data.

Moreover, it was interesting to observe that the two studies which enquired about people's preferences, experienced a smaller volume of disclosure than studies which only presented affirmative motivational messages. These findings suggest that, in contrast to experienced citizen scientists (Bowser et al., 2017), awareness of one's motivations and motivational priorities does not facilitate the sharing of personal data among dabblers.

In future research, it will be important to clarify why this is the case. Perhaps asking citizen scientists to consider their preferences just before disclosure makes them evaluate the project more critically? One solution to this would be to avoid the discussion of personal preferences with dabblers. However, this would go against the values of transparency and cooperation, which are at the heart of citizen science. An alternative solution would be to develop several versions of the same project so that participants can choose one that aligns with their preferences and provides the benefits they value the most. If effective at increasing data disclosure, such an approach could help empower citizen scientists to make their own decisions about the incentives they expect when asked to share personal data within projects.

6.3.3. The differences between citizen science surveys and online quizzes

In chapter 2 we discussed how the negative media attention given to online quizzes (Cadwalladr, 2018) might result in a lower level of data disclosure by people who sign up for an online quiz when compared to those who sign up for a citizen science survey. This pattern was indeed identified in Study 2 – despite filling the same survey, participants who believed that they were taking an online quiz disclosed a smaller volume of data. Why did this happen? It is possible that people who join an online quiz which offers the discovery of a personal score, wish to find out something about themselves while also trying to minimise the disclosure of sensitive data. Although consumer surveys and research into the privacy paradox suggest that people are only concerned about privacy in theory and that privacy concern does not translate into the data sharing choices they make (Kokolakis, 2017) – Sannon et al. (Sannon, Bazarova, & Cosley, 2018) introduce the concept of privacy lies, a common strategy where individuals disclose false information online, as a data protection strategy. Is it, therefore, possible that designing citizen science projects to resemble the online quizzes hosted by private companies may diminish the hosts' ability to gather personal data? Our findings suggest that this might be the case. A lack of emphasis on the concept of citizen science at the point of recruitment may help attract a wider range of participants (opening recruitment up to individuals without a pre-existing interest in citizen science), however, will ultimately hurt the chances of projects gathering the necessary data.

Finally, because the label of 'citizen science' appears to attract a larger volume of data disclosure, it is important to consider whether private companies that conduct for-profit research should be able to call their data collection efforts citizen science, or whether the label should be reserved for university and other public institution-led research.

While the lack of strict boundaries can allow a wider breadth of projects and therefore attract new audiences to the citizen science community, it may also make it difficult to adequately protect both the citizen scientists from bad actors, and the citizen science brand. One solution could be for universities to collectively develop an accrediting system for citizen science projects.

6.3.4. Gender ratios among the participants

Across all five studies reported in this thesis, the large majority of participants stated their gender as female. Overall, women made up 67.4% (708 out of 1050) of all participants. Considering the voluntary nature of the contribution to the surveys, this is consistent with reports that women are often responsible for an array of unpaid tasks (Jarrett, 2014). The predominance of female participants, however, is not a set pattern in citizen science. For example, the astronomy-focused *Galaxy Zoo*, one of the largest citizen science projects, mainly attracts male citizen scientists (Raddick et al., 2013). The gender balance of super-contributors in a humanities-focused citizen science project *Transcribe Bentham* was slightly skewed towards males (Seaward, 2018).

The pattern of participation in Studies 1-5 may suggest that women are more likely to take part in psychological citizen science. Indeed, at universities, psychology proves to be a ‘female-concentrated discipline’ (Marulanda & Radtke, 2019). However, it could also be that women are more likely to take part in citizen science projects that require the disclosure of personal data. In future research, it will be important to examine whether women are more likely to join disclosure-intensive citizen science projects and what factors may account for this. This is vital for attracting diverse populations to citizen science initiatives, but also for protecting overrepresented samples (in this case women) from a disproportionate burden of making disclosure decisions.

6.4. Practical recommendations for project coordinators

6.4.1. Motivation to participate versus motivation to disclose

We found that the motivations that prove most popular and encourage the largest number of participants to complete the survey are not the motivations associated with the largest volume of data disclosure. These findings suggest that motivation to participate and motivation to disclose information are two separate concepts and projects that rely on the disclosure of sensitive data may not necessarily benefit from emphasising the benefits of citizen science that attract the largest number of participants.

When employing motivational messages, it may be useful to pilot a citizen science project with a small sample of participants, to determine which motivational messages result in the largest volume of data disclosure. Such information can then be used to design the project to offer such benefits and incorporate communication about them into the project. At the same time, coordinators of citizen science initiatives that rely on the disclosure of personal data may need to be realistic about the ability to recruit a large number of participants while sustaining necessary levels of data disclosure. In general, in the context of citizen science, one of these goals – recruitment of a large sample or an increase in data disclosure – may need to be prioritised at the expense of the other.

6.4.2. Prompted motivation versus preferred motivation

Our studies have demonstrated that, when participants are presented with short motivational messages, they disclose a larger volume of sensitive data than when their attention is directed to their motivational preferences. These findings suggest that, when aiming to obtain personal data, it is better to affirm existing benefits of participation rather than asking participants to engage in self-reflection. The latter could cause them to assess a project more critically and, as a result, share a smaller volume of data.

The findings from Study 5 suggest that the only group of participants that share a larger volume of data when prompted to consider their motivational preferences are individuals who value social interaction. As the online survey used in this study did not offer opportunities for social interaction, this effect could be due to the interaction already available within the existing online community or the individual characteristics of those who chose the Social interaction option as their answer to the motivational question. It appears that recruiting people who are eager to interact with other participants could result in larger volumes of data disclosure. It should be noted, however, that projects which offer interaction between participants can open new avenues for privacy breaches (for example when sharing personal data on forums meant for discussion about the project (Bowser et al., 2017)). Therefore, the provision of social interaction tools to encourage data disclosure among citizen scientists could also require the provision of cybersecurity training to participants.

6.4.3. The impact of project type

Study 3 demonstrated that participants who join a citizen science project share a larger volume of sensitive data than those who join an online quiz. This finding suggests that the concept of citizen science should be emphasised at the point of recruitment: this will result in a sample of participants who will disclose a larger volume of data. While potentially beneficial for citizen science coordinators, this finding can also pose a risk to the citizen science community, should bad actors use the label of ‘citizen science’ to solicit data online. Such a scenario is not unlikely, considering the proliferation of online quizzes and other devices that aim to convince people to share personal data online. As discussed above, it will become increasingly important for the citizen science community to develop a standardised validation system that protects participants from bad actors.

6.4.4. The importance of transparent communication

The research reported in this thesis was conducted to support transparent and fair communication with citizen scientists when encouraging them to disclose personal data. While communication about the benefits of participation can increase levels of data disclosed within projects, any communication with participants must have a grounding in reality. Participants who are promised a certain outcome and are not satisfied that it has been delivered may decide to withdraw their data from a project. Moreover, the General Data Protection Regulation requires all data collectors and processors to provide honest and reliable information at every stage of soliciting personal data. Citizen science projects should only make the promises they can fulfil to comply with both ethical and regulatory standards.

6.5. The ethics of studying data disclosure in citizen science

To the best of our knowledge, the research presented in this thesis is the first experimental investigation of data disclosure in citizen science. As such, we are building a foundation for research in this area. There are several important ethical considerations that, we believe, need to be recognised when conducting such research and interpreting as well as implementing the findings from studies. Below, we discuss several ethical issues: recruitment of participants from an existing online community, the use of deception in Studies 1-5, addressing complaints from study participants and, crucially, responsible application of the findings from this thesis.

6.5.1. Recruiting from an existing online community

In the case of Study 4 and Study 5, we recruited participants from an online community of pet owners, who are active and interact with one another on

Twitter. It is important to consider the ethical considerations of recruiting participants from an existing community. We believe that it is vital to be mindful of the community's reactions to the research conducted. While we received a small number of complaints (further discussed in section 6.5.3.), we also received much positive feedback, with the members of the community stating that they 'enjoyed the survey', found it to be 'clever' and 'interesting' and wished the researcher luck with completing the study. This feedback suggested that, overall, the community was accepting of the recruitment process.

6.5.2. Use of deception

The studies described in this thesis used deception. Firstly, the participants believed that they were taking part in a real citizen science project that studied the connection between stress and sleep habits (in Studies 1-3) or the connection between stress and pet ownership (in Studies 4 and 5). However, in reality, they were filling fictional surveys that were created specifically to study the disclosure behaviour of citizen scientists.

To emulate the authentic context of citizen science, we created five surveys: 'Sleep Mapping', 'Sleep Experience', 'Sleep Patterns', 'Pet Owners' and 'Pet Experience' (for, respectively, Studies 1, 2, 3, 4 and 5). This element of deception was used for several reasons. Firstly, as noted by Bowser et al. (Bowser et al., 2017), 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 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.

Furthermore, in each study, the participants only became aware that they were taking part in a data disclosure study once they have completed the survey - specifically at the point of reading the debriefing message. This was necessary to ensure that we could study the participants’ authentic disclosure behaviour as opposed to their opinions or attitudes.

We took steps to ensure transparency about the use of deception. The Participant Information Sheet, which all participants were asked to read before filling the survey, explained that many psychological studies involve some level of deception and that the full context and aim of the study would be revealed following the completion of the survey. At the end of the survey, all participants were provided with a debriefing message explaining the aims of the study and outlining the reasons why the use of deception was necessary. Participants were able to withdraw from the study without stating the reason, and any such requests were fulfilled promptly.

In the future, it may be worthwhile to explore ways in which research into the disclosure behaviour of citizen scientists could be conducted without, or with less, deception. It will be important to study the degree to which citizen scientists, as a population, are accepting of research inquiring into their behaviour. It could be helpful to raise awareness among citizen scientists about why some researchers see it as important to study their behaviour. It is likely that the higher the ethical standards of the citizen science project designers and coordinators, the more accepting the participants may be of such research. Namely, if it is understood that findings from research about citizen scientists’ behaviour will only be used in ethical and transparent ways – and not to manipulate participants – the easier it is likely to be to gain acceptance of such research.

6.5.3. Addressing complaints from the participants

Throughout conducting this research, we received feedback from participants, both via social media, as well as via email. Nine individuals wrote to tell us that they viewed the use of deception in research as unacceptable. All of such emails and messages received prompt responses explaining why the use of deception was, in this case, a necessary element of studying authentic disclosure behaviour.

Two participants were concerned that the existence of a citizen science survey that asks for sensitive personal data might normalise such requests in other contexts (for example, they worried that a person who fills our survey might be less cautious when another, less reputable, organisation requests their sensitive data). These comments allowed us to improve and refine the debriefing information, which we provided each participant with. We added a message to the debriefing, that cautioned participants against indiscriminately sharing their data on the Internet, as well as directing them to a website with information about staying safe online.

Throughout the data collection process, we made every effort to remind the participants that taking part in the study was voluntary and that they could withdraw from the process and ask for their data to be deleted if they changed their mind about participation. Any complaints from participants were discussed with the supervisory team to ensure that the participant received an appropriate response and that their feedback would be, where possible, incorporated in the way we conducted the studies.

6.5.4. Ethical use of findings reported in this thesis

This thesis demonstrates practical ways to encourage citizen scientists to disclose sensitive data within projects. It is important to note that the use of motivational messages, with the aim of influencing participant behaviour, can only be considered ethical when these messages communicate accurate and honest information about the project. In practice, this would mean that if a project is advertised as enhancing learning, then steps must be taken to design and execute the project such that it does facilitate acquisition of new knowledge and/or skills. These outcomes should not be assumed and instead should be either assessed empirically or supported by design that is driven by empirical findings.

We hope that the work presented in this thesis will help researchers to design future citizen science projects in a way that is better aligned with people's preferences and motivations related to 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. We propose that findings from studies such as Studies 1-5 must be utilised to design projects that overtly inform participants about any motivational measures used to increase disclosure. We make a clear distinction between covertly coercing participants to disclose data and learning about people's preferences 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.

6.6. Conclusion

In this chapter, we addressed the main research aim and discussed how the findings from this research can be utilised by people who design and/or coordinate citizen science projects. We also discussed the ethical implications of studying data disclosure behaviour in the context of citizen science. While this thesis puts a marked emphasis on encouraging the disclosure of data among citizen scientists, it is important to remember that we can only do this when we are certain that participants’ data will be stored securely and used only for the purposes clearly outlined during consent procedures. Furthermore, while communication about the benefits of taking part in citizen science can be a tool for achieving increased disclosure of data, this tool must be used responsibly. We ought to only make the promises that we can fulfil, and so the use of motivational messages that advertise benefits of participation should be underpinned by project design that can reliably and realistically deliver those benefits. Encouraging the disclosure of personal data in citizen science projects that require such information has to happen through ethical and sustainable means – and the studies reported in this thesis provide a foundation for this.

References

Aspler, S. (2019). Customize Your Cheesecake And We'll Reveal Which High School Clique You Belong In. Buzzfeed Quiz. Retrieved November 27, 2019, from https://www.buzzfeed.com/sarahaspler/bake-a-cheesecake-and-well-reveal-which-high-school-clique?

August, T., Oliviera, N., Tan, C., Smith, N. A., & Reinecke, K. (2018). Framing Effect: Choice of Slogans Used to Advertise Online Experiments Can Boost Recruitment and Lead to Sample Biases. *ACM Hum.-Comput. Interact.2, CSCW*.

Balebako, R., Jung, J., Lu, W., Cranor, L. F., & Nguyen, C. (2013). Little brothers watching you. *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 1. https://doi.org/10.1145/2501604.2501616

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2017.04.013

Bath, P. A. (2008). Health informatics: Current issues and challenges. *Journal of Information Science*, *34*(4), 501–518. https://doi.org/10.1177/0165551508092267

Bisceglio, P. (2017). The Dark Side of That Personality Quiz You Just Took.

Blackmore, C., Tantam, D., & Deurzen, E. van. (2008). Evaluation of e-learning outcomes: experience from an online psychotherapy education programme. *Open Learning: The Journal of Open, Distance and e-Learning*. https://doi.org/10.1080/02680510802420027

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding

Science Knowledge and Scientific Literacy. *BioScience*. https://doi.org/10.1525/bio.2009.59.11.9

Bowser-Livermore, A., & Wiggins, A. (2015). Privacy in Participatory Research: Advancing Policy to support Human Computation. *Human Computation*, *2*(1). https://doi.org/10.15346/hc.v2i1.3

Bowser, A., Shilton, K., Preece, J., & Warrick, E. (2017). Accounting for privacy in citizen science: Ethical research in a context of openness Volunteer Motivation. *Proceedings of CSCW 2017*. https://doi.org/10.1145/2998181.2998305

Cadwalladr, C. (2018). The Cambridge Analytica Files. *The Guardian*.

Caplan, S. E. (2003). Preference for Online Social Interaction: A Theory of Problematic Internet Use and Psychosocial Well-Being. *Communication Research*. https://doi.org/10.1177/0093650203257842

Causer, T., & Wallace, V. (2012). Building A Volunteer Community: Results and Findings from Transcribe Bentham. *Digital Humanities Quarterley*.

Chong, A. Y. L., Ooi, K. B., Lin, B., & Tan, B. I. (2010). Online banking adoption: An empirical analysis. *International Journal of Bank Marketing*. https://doi.org/10.1108/02652321011054963

Cialdini, R. (2001). Harnessing the science of persuasion. *Harvard Business Review*.

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media and Society*. https://doi.org/10.1177/1461444816657096

de la Torre, L. (2018). A Guide to the California Consumer Privacy Act of 2018. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3275571

Diner, D., Nakayama, S., Nov, O., & Porfiri, M. (2018). Social signals as design interventions for enhancing citizen science contributions. *Information,* *Communication & Society*, *21*(4), 594–611.

Domroese, M. C., & Johnson, E. A. (2017). Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation*. https://doi.org/10.1016/j.biocon.2016.08.020

Dunn, S., & Hedges, M. (2012). Crowd-Sourcing Scoping Study Engaging the Crowd with Humanities Research. *Centre for E-Research, King's College London*.

Dutta, P., Aoki, P. M., Kumar, N., Mainwaring, A., Myers, C., Willett, W., & Woodruff, A. (2009). Demo abstract: Common sense - Participatory urban sensing using a network of handheld air quality monitors. *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, SenSys 2009*. https://doi.org/10.1145/1644038.1644095

Eccles, K., & Greg, A. (2014). Your Paintings Tagger: Crowdsourcing Descriptive Metadata for a National Virtual Collection. In M. Ridge (Ed.), *Crowdsourcing our Cultural Heritage* (pp. 185–208). Suurey: Ashgate.

European Union. (2016). Regulation 2016/679 of the European parliament and the Council of the European Union. *Official Journal of the European Communities*. https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf

Eveleigh, A., Jennett, C., Blandford, A., Brohan, P., & Cox, A. L. (2014). Designing for dabblers and deterring drop-outs in citizen science. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*. https://doi.org/10.1145/2556288.2557262

Gómez-Barroso, J. L. (2018). Experiments on personal information disclosure: Past and future avenues. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2018.03.017

Goodwin, C. (2019). Privacy: Recognition of a Consumer Right. *Journal of Public Policy & Marketing*. https://doi.org/10.1177/074391569101000111

Haklay, M. (2015). *Citizen Science and Policy: A European Perspective*. Retrieved from www.CommonsLab.wilsoncenter.org

Harris and Associates Inc., & Westin, A. (1998). *E- commerce and privacy: What net users want.*

Hollandsworth, R., Donovan, J., & Welch, M. (2017). Digital Citizenship: You Can't Go Home Again. *TechTrends*. https://doi.org/10.1007/s11528-017-0190-4

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Internacional Journal of Chronobiology*. https://doi.org/10.1177/0748730405285278

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*. https://doi.org/10.1109/MC.2018.3191268

Jarrett, K. (2014). The relevance of "women's work": Social reproduction and immaterial labor in digital media. *Television and New Media*. https://doi.org/10.1177/1527476413487607

Jennett, C., Furniss, D. J., Iacovides, I., Wiseman, S., Gould, S. J. J., & Cox, A. L. (2014). Exploring Citizen Psych-Science and the Motivations of Errordiary Volunteers. *Human Computation*, *1*(2). https://doi.org/10.15346/hc.v1i2.10

Karwatzki, S., Dytynko, O., Trenz, M., & Veit, D. (2017). Beyond the Personalization–Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization. *Journal of Management Information Systems*, *34*(2), 369–400. https://doi.org/10.1080/07421222.2017.1334467

Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: The effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*. https://doi.org/10.1111/isj.12062

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers and Security*. https://doi.org/10.1016/j.cose.2015.07.002

Kullenberg, C., & Kasperowski, D. (2016). What is citizen science? - A scientometric meta-analysis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0147152

Land-Zandstra, A. M., van Beusekom, M. M., Koppeschaar, C. E., & van den Broek, J. M. (2016). Motivation and learning impact of Dutch flu-trackers. *Journal of Science Communication*. https://doi.org/10.22323/2.15010204

Lee, T. K., Crowston, K., Harandi, M., Østerlund, C., & Miller, G. (2018). Appealing to different motivations in a message to recruit citizen scientists: Results of a field experiment. *Journal of Science Communication*. https://doi.org/10.22323/2.17010202

Leon, S. M. (2014). Build, Analyse, and Generalise: Community Transcription of the Papers of the War Department and the Development of Scripto. In M. Ridge (Ed.), *Crowdsourcing Our Cultural Heritage* (pp. 88–112). Surrey: Ashgate.

Mackert, M., Koh, H. E., Mabry-Flynn, A., Champlin, S., & Beal, A. (2017). Domestic and International College Students: Health Insurance Information Seeking and Use. *Journal of International Students*. https://doi.org/http://dx.doi.org/10.5281/zenodo.570016

Malheiros, M., Brostoff, S., Jennett, C., & Sasse, M. A. (2013). Would you sell your mother’s data? personal data disclosure in a simulated credit card application. In *The Economics of Information Security and Privacy*.

https://doi.org/10.1007/978-3-642-39498-0_11

Malheiros, M., Preibusch, S., & Sasse, M. A. (2013). Fairly truthful: The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-642-38908-5_19

Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. C. (2017). Now that you mention it: A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2017.03.024

Marulanda, D., & Radtke, H. L. (2019). Men Pursuing an Undergraduate Psychology Degree: What's Masculinity Got to Do with It? *Sex Roles*. https://doi.org/10.1007/s11199-018-0995-4

Mothersbaugh, D. L., Foxx, W. K., Beatty, S. E., & Wang, S. (2012). Disclosure Antecedents in an Online Service Context: The Role of Sensitivity of Information. *Journal of Service Research*. https://doi.org/10.1177/1094670511424924

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto, CA: Stanford University Press.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). *The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors*.

Nov, O., Arazy, O., & Anderson, D. (2011). Dusting for science: Motivation and participation of digital citizen science volunteers. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/1940761.1940771

Oberhauser, K., & Lebuhn, G. (2012). Insects and plants: Engaging undergraduates in authentic research through citizen science. *Frontiers in*

Ecology and the Environment. https://doi.org/10.1890/110274

Ofcom. (2015). *Adults' Media Use and Attitudes Report 2015*. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0014/82112/2015_adults_media_use_and_attitudes_report.pdf

Ofcom. (2016). *Adults' Media Use and Attitudes Report 2016*. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0026/80828/2016-adults-media-use-and-attitudes.pdf

Ofcom. (2017). *Adults' Media Use and Attitudes Report 2017*. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0020/102755/adults-media-use-attitudes-2017.pdf

Oosterman, J., Bozzon, A., Houben, G.-J., Nottamkandath, A., Dijkshoorn, C., Aroyo, L., & Traub, M. C. (2014). Crowd vs . Experts : Nichesourcing for Knowledge Intensive Tasks in Cultural Heritage. *WWW'14 Companion*. https://doi.org/10.1145/2567948.2576960

Otterbacher, J. (2015). Crowdsourcing Stereotypes: Linguistic Bias in Metadata Generated via GWAP. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2702123.2702151

Pearce, K. E., & Rice, R. E. (2013). Digital Divides From Access to Activities: Comparing Mobile and Personal Computer Internet Users. *Journal of Communication*. https://doi.org/10.1111/jcom.12045

Phillips, L. B. (2014). The Role of Open Authority in a Collaborative Web. In M. Ridge (Ed.), *Crowdsourcing Our Cultural Heritage* (pp. 247–267). Surrey: Ashgate.

Poikela, M., & Kaiser, F. (2017). *It Is a Topic That Confuses Me - Privacy Perceptions in Usage of Location-Based Applications*.

https://doi.org/10.14722/eurousec.2016.23010

Ponciano, L., Brasileiro, F., Simpson, R., & Smith, A. (2014). Volunteers' engagement in human computation for astronomy projects. *Computing in Science and Engineering*. https://doi.org/10.1109/MCSE.2014.4

Preece, J. (2016). Citizen Science: New Research Challenges for Human–Computer Interaction. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2016.1194153

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murray, P., … Vandenberg, J. (2013). Galaxy zoo: Motivations of citizen scientists. *Astronomy Education Review*. https://doi.org/10.3847/AER2011021

Ridge, M. (2013). From Tagging to Theorizing: Deepening Engagement with Heritage through Crowdsourcing. *Curator: The Museum Journal*, *56*(4), 435–450.

Riesch, H., & Potter, C. (2014). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*. https://doi.org/10.1177/0963662513497324

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., … Jacobs, D. (2012). *Dynamic changes in motivation in collaborative citizen-science projects*. https://doi.org/10.1145/2145204.2145238

Sannon, S., Bazarova, N. N., & Cosley, D. (2018). Privacy lies: Understanding how, when, and why people lie to protect their privacy in multiple online contexts. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3173574.3173626

Schade, S., & Tsinaraki, C. (2016). Survey Report: Data Management in Citizen Science Projects. In *JRC Technical Reports*. https://doi.org/10.2788/539115

Seaward, L. (2018). Project Update – Transcribe Bentham, women and citizen science. Retrieved March 9, 2020, from https://blogs.ucl.ac.uk/transcribe-bentham/2018/03/23/project-update-women-and-citizen-science/

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2009.03.017

Socientize Project. (2013). Green paper on Citizen Science. Citizen Science for Europe: Towards a society of empowered citizens and enhanced research. *Socientize*.

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*. https://doi.org/10.2307/40041279

Steinfeld, N. (2015). Trading with privacy: The price of personal information. *Online Information Review*. https://doi.org/10.1108/OIR-05-2015-0168

Stevens, M., Vitos, M., Altenbuchner, J., Conquest, G., Lewis, J., & Haklay, M. (2014). Taking participatory citizen science to extremes. *IEEE Pervasive Computing*. https://doi.org/10.1109/MPRV.2014.37

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292. https://doi.org/10.1016/J.BIOCON.2009.05.006

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2012.11.022

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., … Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*. https://doi.org/10.1016/j.biocon.2014.10.021

Thomas, S. M., & Jha, R. (2017). How Privileged Are You? Buzzfeed Quiz.

Retrieved November 27, 2019, from https://www.buzzfeed.com/soniathomas/how-privileged-are-you?

TRUSTe. (2016). *U.S. consumer privacy index*. Retrieved from https: / / www . truste . com / resources / privacy - research / ncsa - consumer - privacy-index-us/.

UK Parliament. Data Protection Act, 2018 c. 12. , UK law § (2018).

Vilhelmson, B., Thulin, E., & Elldér, E. (2017). Where does time spent on the Internet come from? Tracing the influence of information and communications technology use on daily activities. *Information Communication and Society*. https://doi.org/10.1080/1369118X.2016.1164741

Ward, S., Bridges, K., & Chitty, B. (2005). Do incentives matter? An examination of on-line privacy concerns and willingness to provide personal and financial information. *Journal of Marketing Communications*. https://doi.org/10.1080/1352726042000263575

Wiseman, S., Cox, A. L., Gould, S. J. J., & Brumby, D. P. (2017). Exploring the effects of non-monetary reimbursement for participants in HCI research. *Human Computation*, *4*(2), 1. https://doi.org/10.15346/hc.v4i1.1

Woodruff, A., Pihur, V., Schmidt, L., Brandimarte, L., & Acquisti, A. (2014). Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. *USENIX Association Tenth Symposium On Usable Privacy and Security*, *1*.

Zuckerberg, M. (2018). *Facebook Post*. Retrieved from https://www.facebook.com/zuck/posts/10104712037900071

Appendix A: The survey

A.1 Participant Information Sheets

A. 2 Consent questions

A.3 Demographic questions

A.4 Neutral Items

A.5 Sensitive Items

A.6 Debriefing messages

A.1. Participant Information Sheets

A.1.1. Participant Information Sheet Question Validation Study

Participant Information Sheet for Healthy Adults in Research Studies

UCL Research Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Neutral and Sensitive Questions in Citizen Science: Two Card Sorting Tasks

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

Why are you conducting this study?

We created two card sorting tasks in order to learn what people think about the design of citizen science surveys. Some surveys ask you to share sensitive data – we want to know which types of survey questions are perceived as requesting sensitive data, and which are perceived as requesting neutral data. Results from this study will help us better understand how citizen scientists perceive the concepts of ‘neutral’ and ‘sensitive’ data – and why they choose to disclose some types of information, while withholding other types of information.

Why have I been chosen to take part?

To take part you must be a healthy adult and aged over 18.

Do I have to take part?

Participation is entirely voluntary. It is up to you to decide whether or not to take part.

What if I change my mind?

You have the right to withdraw from the study at any point up to submission – please note that we cannot delete data following submission as all responses will be anonymous.

What will happen to me if I take part?

You will be asked to complete two short card sorting tasks: this will take you approximately 15 minutes. In each card sorting task, you will be asked to imagine that you are taking part in a citizen science survey. You will be asked to sort survey questions into 4 categories (Definitely Neutral, Somewhat Neutral, Somewhat Sensitive, and Definitely Sensitive) based on whether you think a particular question requests sensitive data.

There are no right or wrong answers – we are interested in your opinions. For example, if you read a question, and think '***This question asks me to share neutral data***', then please assign it to one of the neutral categories. If you read a question and think '***This question asks me to share sensitive data***' then please assign it to one of the sensitive categories.

What are the possible disadvantages or risks of taking part?

There are no foreseeable discomforts, disadvantages or risks of taking part.

What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}. If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

Please CIRCLE one of the following options

I wish to proceed with the study

I do not wish to proceed with the study

____

Full name and date

A.1.2. Participant Information Sheet Study 1

Participant Information Sheet for Healthy Adults in Research Studies

UCL Research Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Sleep Mapping

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

1. Invitation Paragraph

You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project’s purpose?

The aim of this experiment is to learn about how different sources of stress in a person’s life are related to their sleep patterns.

3. Why have I been chosen?

To take part you must be a healthy adult and aged over 18. We aim to test 50+ participants in this study.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to fill a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up to that point.

5. What will happen to me if I take part?

The study duration (including reading this information sheet) is about 20 minutes. It takes place online. First you will be asked to fill a consent form. Then, if you consent to take part in the study, you will be asked to answer 38 short questions.

All data will be collected and stored in accordance with the Data Protection Act 1998.

6. What are the possible disadvantages and risks of taking part?

There are no foreseeable discomforts, disadvantages and risks for taking part. However, some of the questionnaire queries will be related to your family history and events, financial and health information. If you do feel uncomfortable, please feel free to withdraw from the study at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will inform research in the area of health psychology.

8. What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}

If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

11. What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

12. Deception

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).

13. Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at {email address}. UCL's Data Protection Officer is {name} and he can also be contacted at {email address}.

Your personal data will be processed for the purposes outlined in this notice.

The legal basis that would be used to process your personal data will be performance of a task in the public interest.

The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project.

If we are able to anonymise or pseudonymise the personal data you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at {email address}. If you remain unsatisfied, you may wish to contact the Information Commissioner’s Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

14. Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

15. Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

A.1.3. Participant Information Sheet Study 2

Participant Information Sheet for Healthy Adults in Research Studies

UCLIC Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Sleep Experience

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

1. Invitation Paragraph

You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project’s purpose?

The aim of this experiment is to learn about how different sources of stress in a person’s life are related to their sleep patterns.

3. Why have I been chosen?

To take part you must be a healthy adult and aged over 18. We aim to test 100+ participants in this study.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to fill a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up to that point.

5. What will happen to me if I take part?

The study duration (including reading this information sheet) is about 20 minutes. It takes place online. First you will be asked to fill a consent form. Then, if you consent to take part in the study, you will be asked to answer 38 short questions.

All data will be collected and stored in accordance with the Data Protection Act 2018 and the General Data Protection Regulation.

6. What are the possible disadvantages and risks of taking part?

There are no foreseeable discomforts, disadvantages and risks for taking part. However, some of the questionnaire queries will be related to your family history and events, financial and health information. If you do feel uncomfortable, please feel free to withdraw from the study at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will inform research in the area of psychology.

8. What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}

If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

11. What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

12. Deception

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).

13. Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at {email address}. UCL's Data Protection Officer is {name} and he can also be contacted at {email address}.

Your personal data will be processed for the purposes outlined in this notice.

The legal basis that would be used to process your personal data will be performance of a task in the public interest.

The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project.

If we are able to anonymise or pseudonymise the personal data, you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at {email address}. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

14. Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

15. Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

A.1.4. Participant Information Sheet Study 3 – citizen science condition

Participant Information Sheet for Healthy Adults in Research Studies

UCLIC Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Sleep Patterns

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

1. Invitation Paragraph

You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project's purpose?

The aim of this experiment is to learn about how different sources of stress in a person's life are related to their sleep patterns.

3. Why have I been chosen?

To take part you must be a healthy adult and aged over 18. We aim to test 100+ participants in this study.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to fill a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up to that point.

5. What will happen to me if I take part?

The study duration (including reading this information sheet) is about 20 minutes. It takes place online. First you will be asked to fill a consent form. Then, if you consent to take part in the study, you will be asked to answer 37 short questions.

All data will be collected and stored in accordance with the Data Protection Act 2018 and the General Data Protection Regulation.

6. What are the possible disadvantages and risks of taking part?

There are no foreseeable discomforts, disadvantages and risks for taking part. However, some of the questionnaire queries will be related to your family history and events, financial and health information. If you do feel uncomfortable, please feel free to withdraw from the study at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will inform research in the area of psychology.

8. What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}

If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

11. What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

12. Deception

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).

13. Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at {email address}. UCL's Data Protection Officer is {name} and he can also be contacted at {email address}.

Your personal data will be processed for the purposes outlined in this notice.

The legal basis that would be used to process your personal data will be performance of a task in the public interest.

The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project.

If we are able to anonymise or pseudonymise the personal data you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at {email address}. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

14. Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

15. Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

A.1.5. Participant Information Sheet Study 3 – online quiz condition

Participant Information Sheet for Healthy Adults in Research Studies

UCLIC Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Sleep Patterns

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

1. Invitation Paragraph

You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project’s purpose?

The aim of this experiment is to learn about how different sources of stress in a person’s life are related to their sleep patterns.

3. Why have I been chosen?

To take part you must be a healthy adult and aged over 18. We aim to test 100+ participants in this study.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to fill a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up to that point.

5. What will happen to me if I take part?

The study duration (including reading this information sheet) is about 20 minutes. It takes place online. First you will be asked to fill a consent form. Then, if you consent to take part in the study, you will be asked to answer 37 short questions.

All data will be collected and stored in accordance with the Data Protection Act 2018 and the General Data Protection Regulation.

6. What are the possible disadvantages and risks of taking part?

There are no foreseeable discomforts, disadvantages and risks for taking part. However, some of the questionnaire queries will be related to your family history and events, financial and health information. If you do feel uncomfortable, please feel free to withdraw from the study at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will inform research in the area of psychology.

8. What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}

If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

11. What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

12. Deception

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).

13. Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at {email address}. UCL's Data Protection Officer is {name} and he can also be contacted at {email address}.

Your personal data will be processed for the purposes outlined in this notice.

The legal basis that would be used to process your personal data will be performance of a task in the public interest.

The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project.

If we are able to anonymise or pseudonymise the personal data you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at {email address}. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

14. Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

15. Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

A.1.6. Participant Information Sheet Study 4

Participant Information Sheet for Healthy Adults in Research Studies

UCL Research Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Pet Experience

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

1. Invitation Paragraph

You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project's purpose?

The aim of this experiment is to learn about how different sources of stress in a person's life are related to their experience of pet ownership.

3. Why have I been chosen?

To take part you must be a healthy adult and aged over 18. We aim to test 180+ participants in this study.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to fill a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up to that point.

5. What will happen to me if I take part?

The study duration (including reading this information sheet) is about 15 minutes. It takes place online. First you will be asked to fill a consent form. Then, if you consent to take part in the study, you will be asked to answer 32 short questions.

All data will be collected and stored in accordance with the Data Protection Act 2018 and the General Data Protection Regulation.

6. What are the possible disadvantages and risks of taking part?

There are no foreseeable discomforts, disadvantages and risks for taking part. However, some of the questionnaire queries will be related to your family history and events, financial and health information. If you do feel uncomfortable, please feel free to withdraw from the study at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for people participating in the project, it is hoped that this work will inform research in the area of psychology.

8. What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}

If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

11. What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

12. Deception

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).

13. Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at {email address}. UCL's Data Protection Officer is {name} and he can also be contacted at {email address}.

Your personal data will be processed for the purposes outlined in this notice.

The legal basis that would be used to process your personal data will be performance of a task in the public interest.

The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project.

If we are able to anonymise or pseudonymise the personal data, you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at {email address}. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

14. Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

15. Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

A.1.7. Participant Information Sheet Study 5

Participant Information Sheet for Healthy Adults in Research Studies

UCL Research Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka.

Title of Study: Pet Owners

Department: UCLIC

Name and Contact Details of the Researcher(s): Anna Rudnicka, UCLIC, University College London, London WC1E 6BT, United Kingdom, {email address}

Name and Contact Details of the Principal Researcher: Prof. Anna L. Cox, UCLIC, University College London, London WC1E 6BT, United Kingdom {phone number}

1. Invitation Paragraph

You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project's purpose?

The aim of this experiment is to learn about how different sources of stress in a person's life are related to their experience of pet ownership.

3. Why have I been chosen?

To take part you must be a healthy adult and aged over 18. We aim to test 100+ participants in this study.

4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to fill a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up to that point.

5. What will happen to me if I take part?

The study duration (including reading this information sheet) is about 15 minutes. It takes place online. First you will be asked to fill a consent form. Then, if you consent to take part in the study, you will be asked to answer 32 short questions.

All data will be collected and stored in accordance with the Data Protection Act 2018 and the General Data Protection Regulation.

6. What are the possible disadvantages and risks of taking part?

There are no foreseeable discomforts, disadvantages and risks for taking part. However, some of the questionnaire queries will be related to your family history and events, financial and health information. If you do feel uncomfortable, please feel free to withdraw from the study at any time.

7. What are the possible benefits of taking part?

Whilst there are no immediate benefits for people participating in the project, it is hoped that this work will inform research in the area of psychology.

8. What if something goes wrong?

If you would like to raise a complaint about this research, please contact the Principal Researcher, Prof. Anna L. Cox, at {email address}

If you feel like your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee – {email address}

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

10. Limits to confidentiality

Confidentiality will be respected subject to legal constraints and professional guidelines.

11. What will happen to the results of the research project?

Results will be disseminated in standard academic outlets. Results may also be disseminated via general interest magazines / newspapers / journals. You will not be identifiable in any report or publication.

12. Deception

Research designs often require that the full intent of the study not be explained prior to participation. Although we have described the general nature of the tasks that you will be asked to perform, the full intent of the study will not be explained to you until after the completion of the study (at which point you may withdraw your data from the study if you wish to do so).

13. Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data and can be contacted at {email address}. UCL's Data Protection Officer is {name} and he can also be contacted at {email address}.

Your personal data will be processed for the purposes outlined in this notice.

The legal basis that would be used to process your personal data will be performance of a task in the public interest.

The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project.

If we are able to anonymise or pseudonymise the personal data you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at {email address}. If you remain unsatisfied, you may wish to contact the Information Commissioner’s Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

14. Who is organising and funding the research?

This research is conducted as part of a PhD grant awarded by the Engineering and Physical Sciences Research Council (EPSRC).

15. Contact for further information

Prof. Anna L. Cox

UCLIC, University College London, London WC1E 6BT, United Kingdom

{phone number}

{email address}

Thank you for reading this information sheet and for considering taking part in this research study.

A.2. Consent Questions

Main consent question

- ○ I wish to proceed with the study
- ○ I do not wish to proceed with the study

Optional questions about future contact

1. I understand that the information I have submitted will be published as a report and I wish to receive a copy of this report.
2. If you would like your contact details to be retained so that you can be contacted in the future by UCL researchers who would like to invite you to participate in follow up studies to this project, or in future studies of a similar nature, please tick the appropriate box below.

Compulsory secondary consent questions

Please note that in order to take part in this study, it is required you answer 'Yes' to the next 15 questions. If you answer 'No' to any of the following 15 questions, you will be redirected to end of the survey and your data will NOT be used in this study.

3. I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me.
4. I consent to take part in an online questionnaire that will ask questions related to my health, sleep behaviour, family history and other personal characteristics and habits.

5. I understand that I will be able to withdraw my data after taking part in the study.
6. I consent to the processing of my personal information (name, surname, age, email and phone number – stored separately from my questionnaire responses and not linked to my questionnaire responses in any other way than through the unique participant number that I will be assigned) for the purposes explained to me. I understand that such information will be handled in accordance with all applicable data protection legislation.
7. I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified.
8. I understand that my information may be subject to review by responsible individuals from the University to include the Engineering and Physical Sciences Research Council for monitoring and audit purposes.
9. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason. I understand that if I decide to withdraw, any personal data I have provided up to that point will be deleted unless I agree otherwise.
10. I understand the potential risks of participating and the support that will be available to me should I become distressed during the course of the research.
11. I understand that the data will not be made available to any commercial organisations but is solely the responsibility of the researcher(s) undertaking this study.
12. I agree that my pseudonymised research data may be used by others for future research. (No one will be able to identify you when this data is shared.)
13. I hereby confirm that I understand the inclusion criteria as detailed in the Information Sheet (above).

14. I hereby confirm that I meet the inclusion criteria (I am a healthy adult and aged over 18).
15. I am aware of who I should contact if I wish to lodge a complaint.
16. I voluntarily agree to take part in this study.
17. I would be happy for the data I provide to be archived at UCLIC. I understand that other authenticated researchers will have access to my pseudonymised data.

A.3. Demographics Questions

1. **How old are you?** (please state in years)
2. **Please state your gender.** (Options: Male; Female; Non-binary; I prefer not to answer this question; I prefer to answer this question in my own words)
3. **How often do you use the Internet?** (Options: All the time; Several times per day; Most days; Several times per week; Less than once a week; Less than once a month)
4. **Have you ever taken part in a Citizen Science project before?** (Options: Yes, once; Yes, more than once a month (please state how many projects you took part in); No; I'm not sure)

A.4. Neutral Items (Morningness-Eveningness Questionnaire)

Adapted from (Horne & Ostberg, 1976).

1. **Considering only your own 'feeling best' rhythm, at what time would you get up if you were entirely free to plan your day?** (Options: 5:00-6:30 a.m.; 6:30-7:45 a.m.; 7:45-9:45 a.m.; 9:45-11:00 a.m.; 11:00 a.m.-12:00 p.m.)
2. **Considering only your own 'feeling best' rhythm, at what time would you go to bed if you were entirely free to plan your evening?** (Options: 8:00-9:00 p.m.; 9:00-10:15 p.m.; 10:15 p.m.-12:30 a.m.; 12:30-1.45 a.m.; 1.45-3.00 a.m.)
3. **If there is a specific time at which you have to get up in the morning, to what extent are you dependent on being woken up by an alarm clock?** (Options: Not at all dependent; Slightly dependent; Fairly dependent; Very dependent)
4. **Assuming adequate environmental conditions, how easy do you find getting up in the mornings?** (Not at all easy; Not very easy; Fairly easy; Very easy)
5. **How alert do you feel during the first half hour after having woken in the mornings?** (Options: Not at all alert; Slightly alert; Fairly alert; Very alert)
6. **How is your appetite during the first half-hour after having woken in the mornings?** (Options: Very poor; Fairly poor; Fairly good; Very good)
7. **During the first half-hour after having woken in the morning, how tired do you feel?** (Options: Very tired; Fairly tired; Fairly refreshed; Very refreshed)

8. **When you have no commitments the next day, at what time do you go to bed compared to your usual bedtime?** (Options: Seldom or never later; Less than one hour later; 1-2 hours later; More than two hours later)
9. **You have decided to engage in some physical exercise. A friend suggests that you do this one hour a week and the best time for him is between 7:00-8:00 a.m. Bearing in mind nothing else but your own 'feeling best' rhythm, how do you think you would perform?** (Options: Would be on good form; Would be on reasonable form; Would find it difficult; Would find it very difficult)
10. **At what time in the evening do you feel tired and as a result in need of sleep?** (Options: 8:00-9:00 p.m.; 9:00-10:15 p.m.; 10:15 p.m. - 12:45 a.m.; 12:45-2:00 a.m.; 2:00-3:00 a.m.)
11. **You wish to be at your peak performance for a test which you know is going to be mentally exhausting and lasting for two hours. You are entirely free to plan your day and considering only your own 'feeling best' rhythm which ONE of the four testing times would you choose?** (Options: 8:00-10:00 a.m.; 11:00 a.m.-1:00 p.m.; 3:00-5:00 p.m.; 7:00-9:00 p.m.)
12. **If you went to bed at 11 p.m. at what level of tiredness would you be?** (Options: Not at all tired; A little tired; Fairly tired; Very tired)
13. **For some reason you have gone to bed several hours later than usual, but there is no need to get up at any particular time the next morning. Which ONE of the following events are you most likely to experience?** (Options: Will wake up at usual time and will NOT fall asleep; Will wake up at usual time and will doze thereafter; Will wake up at usual time but will fall asleep again; Will NOT wake up until later than usual)

14. **One night you have to remain awake between 4-6 a.m. in order to carry out a night watch. You have no commitments the next day, which ONE of the following alternatives will suit you best?** (Options: Would NOT go to bed until watch was over; Would take a nap before and sleep after; Would take a good sleep before and a nap after; Would take ALL sleep before watch)

15. **You have to do two hours of hard physical work. You are entirely free to plan your day and considering only your own 'feeling best' rhythm which ONE of the following times would you choose?** (Options: 8:00-10:00 a.m.; 11:00 a.m.-1:00 p.m.; 3:00-5:00 p.m.; 7:00-9:00 p.m.)

16. **You have to engage in hard physical exercise. A friend suggests that you do this for one hour twice a week and the best time for him is between 10-11 p.m. Bearing in mind nothing else but your own 'feeling best' rhythm how well do you think you would perform?** (Options: Would be on good form; Would be on reasonable form; Would find it difficult; Would find it very difficult)

17. **Suppose that you can choose your own work hours. Assume that you worked a FIVE-hour day (including breaks) and that your job was interesting and paid by results. Which FIVE CONSECUTIVE HOURS would you select?** (Options: 3.00-7.00 a.m.; 7.00-9.00 a.m.; 9.00.a.m. - 1.00.p.m.; 1.00-4.00p.m.; 4.00p.m. - 3.00.a.m.)

18. **At what time of the day do you think that you reach your 'feeling best' peak?** (Options: 4:00-7:00 a.m.; 7:00-9:00 a.m.; 9:00 a.m. - 4:00 p.m.; 4:00-9:00 p.m.; 9:00 p.m. - 4:00 a.m.)

19. **One hears about 'morning' and 'evening' types of people. Which ONE of these types do you consider yourself to be?** (Options:

Definitely a 'morning' type; Definitely more a 'morning' than an evening type; Rather more an 'evening' type than a 'morning' type; Definitely an 'evening' type)

A.5. Neutral Items (Pet owners' questionnaire)

1. **Did you own a pet when you were a child? Please specify the species of your childhood pet.** (Options: Yes; No)
2. **Does your pet have any health issues?** Please list... (Options: Yes; No)
3. **Does the pet you currently own live with you?** (Options: Yes; No)
4. **Could you list the 3 foods your pet enjoys most?** (Options: Yes; I prefer not to)
5. **Do you have more than one pet?** (Options: Yes; No)
6. **Are you planning to have more pets in the future?** (Options: Yes; No)
7. **Do you buy birthday gifts for your pet?** (Options: Yes; No)
8. **Please tell us the exact age of your pet.** (Options: My pet is … old; I prefer not to answer this question)
9. **What is your pet's favourite toy? Please describe...** (Options: My pet's favourite toy is: …; I prefer not to answer this question)
10. **What are the 3 words that most accurately describe your pet? Please list...** (Options: The three words that most accurately describe my pet are: …; I prefer not to answer this question)
11. **What are the 3 things that most annoy your pet? Please list...** (Options: The 3 things that most annoy my pet are: …; I prefer not to answer this question)
12. **Does your pet ever wear a body warmer or any other type of pet clothing?** (Options: Yes; No)
13. **How many hours does your pet spend napping during the day?** (Options: The number of hours my pet naps during the day is: …; I prefer not to answer this question)

14. **Do you think that owning a pet nowadays is expensive?** (Options: Yes; No)

A.6. Sensitive Items

Adapted from (Malheiros, Brostoff, et al., 2013).

1. **Did any of your loved ones die while you were growing up? Please give their relation to you.** (e.g. mother, brother, friend, etc.). (Options: Yes; No)
2. **Do you suffer from any medical conditions? Please list...** (Options: Yes, No)
3. **Did you live with both your mother and father while you were growing up?** (Options: Yes, No)
4. **Could you list the names and either phone numbers or email addresses of three of your closest friends?** (Options: Yes, I prefer not to)
5. **Do you give us permission to contact your local council to get a copy of your council tax payment history?** (Options: Yes, No)
6. **Do you give us permission to obtain a copy of your TV licence payment history?** (Options: Yes, No)
7. **Do you give us permission to obtain a copy of your gas or electricity payment history?** (Options: Yes, No)
8. **Please provide the name and address (or other contact details) of a previous employer so that we can request a copy of the last recommendation from him/her about you…** (Options: contact details: …; I prefer not to answer this question)
9. **What is the job of your partner / spouse? Please describe…** (Options: The job of my partner / spouse is: …; I prefer not to answer this question)
10. **What are the names of 3 people that you are friends with on a social networking site (facebook, twitter) whose profiles you would be**

happy to share with us? Please list… (Options: The names of 3 people that I am friends on a social networking site and whose profiles I would be happy to share with you are: …; I prefer not to answer this question)

11. **What are the names of 3 people that you are friends with on a professional networking site (LinkedIn, Orkut) whose profiles you would be happy to share with us? Please list...** (Options: The names of 3 people that I am friends on a professional networking site and whose profiles I would be happy to share with you are: …; I prefer not to answer this question)
12. **Will you allow us to measure the typical number and length of messages between you and your friends on social networking sites?** (Options: Yes; No)
13. **What is the length of the longest relationship you have had with a partner / spouse? (years/ months/ weeks)?** (Options: The length of the longest relationship I have had with a partner / spouse is: …; I prefer not to answer this question)
14. **May we obtain a copy of your insurance claims (e.g. car, house)?** (Options: Yes, No)

A.7. Debriefing messages

A.7.1. Debriefing message Study 1

EXPERIMENT DEBRIEF INFORMATION

‘INFORMATION DISCLOSURE IN CITIZEN SCIENCE’

Dear Participant,

Thank you for completing this survey. We would like to give you more information about the purpose and design of this study. This survey served two goals. Firstly, it helped us gather information about how individuals answer questions about their circadian rhythms (for example whether you are the ‘morning type’ or the ‘evening type’). Secondly, we wanted to learn about how different individuals respond to queries about very personal information concerning themselves and their friends or family.

We aim 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. Such research is of particular importance following the recent disclosure of ‘breach of trust’ and potential misuse of Facebook data. If we are to help design safeguards that protect individuals from their data being compromised, we must first study the specifics of disclosure behaviour, that is, when, how, and why individuals make decisions to part with their, often very personal, data. We hope that this work will help researchers to design citizen science projects in future in a way that is more aligned with people’s attitudes and preferences in relation to data privacy and data disclosure.

Before starting the survey, each participant was presented with one of the four, randomly assigned, messages: 1. ‘Extend your knowledge of health psychology by participating in the Sleep Mapping survey!’; 2. ‘Join your fellow citizen scientists in establishing connections between stress and sleep!... Many citizen scientists are already participating in the project…'; 3. 'You can contribute to science by answering questions about sources of stress in your life and quality of sleep!'; 4. 'Health psychology needs your help to connect sources of stress to patterns of sleep behaviour!'. After consenting to take part, each participant was asked 19 questions relating to sleep behaviour – answers to these non-sensitive queries will be used as a 'baseline' to compare with the patterns of response to the next 14 questions about various sources of life stress (sensitive questions). Finally, participants were presented with the single-question privacy scale developed by Westin, which allows us to classify individuals into *Privacy Pragmatists*, *Privacy Fundamentalists*, and *Privacy Unconcerned*.

We are interested in whether the type of message presented to participants, and/or their classification according to Westin's scale, could explain the patterns of information disclosure throughout the rest of the survey. Specifically, we set out to test four hypotheses:

Hypothesis 1: Sensitivity of a data item will be correlated with the proportion of participants willing to respond to it.

Hypothesis 2: *Privacy Fundamentalists* will be less willing to disclose information.

Hypothesis 3: There will be an interaction between type of 'motivational message' and the impact of 'item sensitivity' on disclosure behaviour.

Hypothesis 4: There will be an interaction between type of 'motivational message' and the impact of the score on Westin's privacy scale on disclosure behaviour.

To conduct this study, we could not tell you in advance that we are studying privacy as that could have made you self-conscious about how you answer questions and eliminated the possibility of studying spontaneous and authentic data disclosure. These are important issues and as outlined already, it is important to understand people's behaviour so that we can hope create fairer disclosure environments for future generations of citizen scientists, as well as building greater data privacy literacy among the wider population.

We want to assure you, however, that the purpose of these deeply personal queries - this refers to questions from 8.1. ‘Did any of your loved ones die while you were growing up?’ through to question 8.14. ‘May we obtain a copy of your insurance claims (e.g., car, house)?’ - was solely to learn WHETHER you decide to respond. This means that after recording whether you gave a response or chose not to answer a particular question, any actual information you gave in response to questions 8.1.-8.14., will be deleted. Such information will not be held nor processed.

Our study is similar to one that has already been published that explored what data people were prepared to disclose in the context of a credit card application. More information on that prior work can be found at [Would You Sell Your Mother’s Data? Personal Data Disclosure in a Simulated Credit Card Application. - UCL Discovery] (http://discovery.ucl.ac.uk/1349987/). The motivational messages were adapted from a previous study on recruitment messages in citizen science and the original text can be found at [Recruiting messages matter: Message strategies to attract citizen scientists | Kevin Crowston] (https://crowston.syr.edu/node/669).

We are thankful for your participation. Your survey responses will be stored securely and separately from any personal identifiers you may have disclosed, such as email address, and we will ensure that all the requirements of the Data Protection Act 1998 are complied with.

If there is any reason why you have changed your mind about being part of this project, please contact us at {email address}, and we will remove your data immediately. You do not have to state a reason for withdrawing from the study. If you have any other questions, thoughts or comments please do not hesitate to contact us. We welcome and are grateful for all feedback.

The lead supervisor for this study is Professor Anna L. Cox, UCLIC, University College London, 66 - 72 Gower Street, London, WC1E 6EA; telephone: {phone number}; email: {email address}.

A.7.2. Debriefing message Study 2

Dear Participant,

You have just completed the Citizen Science survey hosted by UCL Interaction Centre (a research centre that is part of University College London). If you have any questions or you changed your mind and would like your data to be removed from the survey then please email us at {email address} (you do not have to state a reason) and we will delete your response.

Please remember that not every individual who asks you for personal data (online, over the phone or in person) can be trusted. **Please always be very cautious when disclosing sensitive personal information about yourself or others (for example information about your health, money or home).** You can read about staying safe online at http://www.bbc.co.uk/webwise/0/21259413

You can now check your sleep score (scroll down to view the sleep score calculated based on your answers to the sleep-related questions in this survey).

Below, we provide more information about the aims and methodology of this survey:

EXPERIMENT DEBRIEF INFORMATION

‘INFORMATION DISCLOSURE IN CITIZEN SCIENCE’

Thank you for completing this survey. We would like to give you more information about the purpose and design of this study. Through this survey, we wanted to learn about how different individuals respond to questions about very personal information concerning themselves and their friends or family. Additionally, participants were able to find out their ‘sleep score’ (which was related to whether the person is more of an ‘early bird’ or more of a ‘night owl’).

We aim 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. Such research is of particular importance following the recent disclosure of ‘breach of trust’ and potential misuse of Facebook data. If we are to help design safeguards that protect individuals from their data being compromised, we must first study the specifics of disclosure behaviour, that is, when, how, and why individuals make decisions to part with their, often very personal, data. We hope that this work will help researchers to design citizen science projects in future in a way that is more aligned with people’s attitudes and preferences in relation to data privacy and data disclosure.

Before starting the survey, each participant was presented with one of the three, randomly assigned, messages: 1) 'Extend your knowledge of health psychology by participating in the Sleep Experience survey!'; 2) 'Fill the Sleep Experience survey and discover your sleep score!'; 3) 'Welcome to the Sleep Experience survey!'. After consenting to take part in the survey, each participant was asked to answer a question about which motivation to participate in citizen science is most important to them: 1. Learning about science, 2. Social interaction, 3. Making a contribution to a project, or 4. Helping scientists. Then, participants were asked to answer 4 Demographics-related questions (enquiring about age, gender, use of Internet and previous citizen science participation), as well as 19 Neutral Items and 14 Sensitive Items. The Neutral Items were questions about sleep habits. The Sensitive Items were questions about different sources of life, relationship and financial stress.

In this study we set out to test three hypotheses. Firstly, we hypothesised that there will be a difference in how much sensitive data people disclose depending on which one of the three motivational messages they saw at the beginning of the survey. Secondly, there will be a difference in how much sensitive data people disclose, depending on which benefits of citizen science are most important to them. Thirdly, we predict that the two factors described above will interact to affect the disclosure of sensitive data.

To conduct this study, we could not tell you in advance that we are studying privacy as that could have made you self-conscious about how you answer questions and eliminated the possibility of studying spontaneous and authentic data disclosure. These are important issues and as outlined already, it is important to understand people's behaviour so that we can hope to create fairer disclosure environments for future generations of citizen scientists, as well as building greater data privacy literacy among the wider population.

We want to assure you, however, that the purpose of these deeply personal queries - this refers to questions from 9.1. 'Did any of your loved ones die while you were growing up?' through to question 9.14. 'May we obtain a copy of your insurance claims (e.g., car, house)?' - was solely to learn WHETHER you decide to respond. This means that after recording whether you gave a response or chose not to answer a particular question, any actual information you gave in response to questions 9.1.-9.14., will be deleted. Such information will not be held nor processed.

Our study is similar to one that has already been published that explored what data people were prepared to disclose in the context of a credit card application. More information on that prior work can be found at [Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card Application. - UCL Discovery] (http://discovery.ucl.ac.uk/1349987/). The motivational messages were adapted from a previous study on recruitment messages in citizen science and the original text can be found at [Recruiting messages matter: Message strategies to attract citizen scientists | Kevin Crowston] (https://crowston.syr.edu/node/669).

We are thankful for your participation. Your survey responses will be stored securely and separately from any personal identifiers you may have disclosed, such as email address, and we will ensure that all the requirements of the Data Protection Act 2018 and the General Data Protection Regulation are complied with.

If there is any reason why you have changed your mind about being part of this project, please contact us at {email address}, and we will remove your data immediately. You do not have to state a reason for withdrawing from the study. If you have any other questions, thoughts or comments please do not hesitate to contact us. We welcome and are grateful for all feedback.

The lead supervisor for this study is Professor Anna L. Cox, UCLIC, University College London, 66 - 72 Gower Street, London, WC1E 6EA; telephone: {phone number}; email: {email address}.

Your sleep score (the number score is visible below, on the left hand side of the percentage score) was calculated based on your answers to the sleep-related questions in this survey. You can check which sleep type you are, based on your score:

70-86 - Definitely Morning Type

59-69 - Moderately Morning Type

42-58 - Neither Type

31-41 - Moderately Evening Type

16-30 - Definitely Evening Type

Please note that this score is based on the closest approximation of the Morningness-Eveningness Questionnaire (Horne & Ostberg, 1976) that was compatible with the online format of this survey - and is **NOT intended as medical advice**. **If you have trouble sleeping, please talk to your doctor, or other qualified healthcare professional.**

You can read about healthy sleep habits at https://www.nhs.uk/live-well/sleep-and-tiredness/how-to-get-to-sleep. Moreover, if you have trouble sleeping, the NHS website offers advice on insomnia and a sleep self-assessment at https://www.nhs.uk/conditions/insomnia.

A.7.3. Debriefing message Study 3

Dear Participant,

You have just completed the Citizen Science survey hosted by UCL Interaction Centre (a research centre that is part of University College London). If you have any questions or you changed your mind and would like your data to be removed from the survey then please email us at {email address} (you do not have to state a reason) and we will delete your response.

Please remember that not every individual who asks you for personal data (online, over the phone or in person) can be trusted. **Please always be very cautious when disclosing sensitive personal information about yourself or others (for example information about your health, money or home).** You can read about staying safe online at http://www.bbc.co.uk/webwise/0/21259413

You can now check your sleep score (scroll down to view the sleep score calculated based on your answers to the sleep-related questions in this survey).

Below, we provide more information about the aims and methodology of this survey:

EXPERIMENT DEBRIEF INFORMATION
‘INFORMATION DISCLOSURE IN CITIZEN SCIENCE’

Thank you for completing this survey. We would like to give you more information about the purpose and design of this study. Through this survey, we wanted to learn about how different individuals respond to questions about very personal information concerning themselves and their friends or family. Additionally, participants were able to find out their ‘sleep score’ (which was related to whether the person is more of an ‘early bird’ or more of a ‘night owl’).

We aim 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. Such research is of particular importance following the recent disclosure of ‘breach of trust’ and potential misuse of Facebook data. If we are to help design safeguards that protect individuals from their data being compromised, we must first study the specifics of disclosure behaviour, that is, when, how, and why individuals make decisions to part with their, often very personal, data. We hope that this work will help researchers to design citizen science projects in future in a way that is more aligned with people’s attitudes and preferences in relation to data privacy and data disclosure.

This study recruits participants for two distinct groups. The first stage involves tweeting messages encouraging people to take part in a citizen science project (‘citizen science’ group). The second stage involves tweeting messages that encourage people to find out their sleep score (‘online quiz’ group).

Before starting the survey, each participant was presented with one of the three, randomly assigned, messages: 1) 'Extend your knowledge of health psychology by participating in the Sleep Patterns survey!'; 2) 'Fill the Sleep Patterns survey and discover your sleep score!'; 3) 'Welcome to the Sleep Patterns survey!'. After consenting to take part in the survey, participants were asked to answer 4 Demographics-related questions (enquiring about age, gender, use of Internet and previous citizen science participation), as well as 19 Neutral Items and 14 Sensitive Items. The Neutral Items were questions about sleep habits. The Sensitive Items were questions about different sources of life, relationship and financial stress.

In this study we aim to explore the differences between the way people disclose personal data in citizen science settings versus in the context of online quizzes. For example, we anticipate that people who are focused on learning their personal score will disclose a smaller volume of data than those who wish to contribute to a citizen science project.

To conduct this study, we could not tell you in advance that we are studying privacy as that could have made you self-conscious about how you answer questions and eliminated the possibility of studying spontaneous and authentic data disclosure. These are important issues and as outlined already, it is important to understand people's behaviour so that we can hope to create fairer disclosure environments for future generations of citizen scientists, as well as building greater data privacy literacy among the wider population.

We want to assure you, however, that the purpose of these deeply personal queries - this refers to questions from 8.1. 'Did any of your loved ones die while you were growing up?' through to question 8.14. 'May we obtain a copy of your insurance claims (e.g., car, house)?' - was solely to learn WHETHER you decide to respond. This means that after recording whether you gave a response or chose not to answer a particular question, any actual information you gave in response to questions 8.1.-8.14., will be deleted. Such information will not be held nor processed.

Our study is similar to one that has already been published that explored what data people were prepared to disclose in the context of a credit card application. More information on that prior work can be found at [Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card Application. - UCL Discovery] (http://discovery.ucl.ac.uk/1349987/).

We are thankful for your participation. Your survey responses will be stored securely and separately from any personal identifiers you may have disclosed, such as email address, and we will ensure that all the requirements of the Data Protection Act 2018 and the General Data Protection Regulation are complied with.

If there is any reason why you have changed your mind about being part of this project, please contact us at {email address}, and we will remove your data immediately. You do not have to state a reason for withdrawing from the study. If you have any other questions, thoughts or comments please do not hesitate to contact us. We welcome and are grateful for all feedback. The lead supervisor for this study is Professor Anna L. Cox, UCLIC, University College London, 66 - 72 Gower Street, London, WC1E 6EA; telephone: {phone number}; email: {email address}.

Your sleep score (the number score is visible below, on the left-hand side of the percentage score) was calculated based on your answers to the sleep-related questions in this survey. You can check which sleep type you are, based on your score:

70-86 - Definitely Morning Type
59-69 - Moderately Morning Type
42-58 - Neither Type
31-41 - Moderately Evening Type
16-30 - Definitely Evening Type

Please note that this score is based on the closest approximation of the Morningness-Eveningness Questionnaire (Horne & Ostberg, 1976) that was compatible with the online format of this survey - and is **NOT intended as medical advice. If you have trouble sleeping, please talk to your doctor, or other qualified healthcare professional**.

You can read about healthy sleep habits at https://www.nhs.uk/live-well/sleep-and-tiredness/how-to-get-to-sleep. Moreover, if you have trouble sleeping, the NHS website offers advice on insomnia and a sleep self-assessment at https://www.nhs.uk/conditions/insomnia.

A.7.4. Debriefing message Study 4

Dear Participant,

You have just completed the Citizen Science survey hosted by UCL Interaction Centre (a research centre that is part of University College London). If you have any questions or you changed your mind and would like your data to be removed from the survey then please email us at {email address} (you do not have to state a reason) and we will delete your response.

Please remember that not every individual who asks you for personal data (online, over the phone or in person) can be trusted. **Please always be very cautious when disclosing sensitive personal information about yourself or others (for example information about your health, money or home).** You can read about staying safe online at http://www.bbc.co.uk/webwise/0/21259413

Below, we provide more information about the aims and methodology of this survey:

EXPERIMENT DEBRIEF INFORMATION

‘INFORMATION DISCLOSURE IN CITIZEN SCIENCE’

Thank you for completing this survey. We would like to give you more information about the purpose and design of this study. This survey served two goals. Firstly, it helped us gather information about how individuals answer questions about their experience of pet ownership. Secondly, we wanted to learn about how different individuals respond to questions about very personal information concerning themselves and their friends or family.

We aim to clarify whether the motivations that drive people to participate in citizen science can also influence how much personal information people will share while taking part in citizen science projects. Such research is of particular importance following the recent disclosure of ‘breach of trust’ and potential misuse of Facebook data. If we are to help design safeguards that protect individuals from their data being compromised, we must first understand when, how, and why individuals make decisions to share personal data. We hope that this work will help researchers to design citizen science projects in future in a way that is more aligned with people’s attitudes and preferences in relation to data privacy and data disclosure.

Before starting the survey, each participant was presented with one of the 4 randomly assigned messages:

1. ‘Extend your knowledge of health psychology by participating in the Pet Owners survey!’ (Learning motivation);
2. ‘Join your fellow citizen scientists in establishing connections between stress and pet ownership behaviours!... Many citizen scientists are already participating in the project…’ (Social motivation);
3. ‘You can contribute to science by answering questions about sources of stress in your life and pet ownership behaviours!’ (Contribution motivation);
4. ‘Health psychology needs your help to connect sources of stress to pet ownership behaviours! (Altruism motivation).

After consenting to take part in the survey, participants were asked to answer 4 Demographics-related questions (enquiring about age, gender, use of Internet and previous citizen science participation), as well as 14 Neutral Items and 14 Sensitive Items. The Neutral Items were questions about the experience of having a pet. The Sensitive Items were questions about different sources of life, relationship and financial stress.

In this study we set out to test two hypotheses. The first hypothesis assumed that people will be more keen to share information when answering neutral questions than when answering sensitive questions. Secondly, we hypothesised that the motivational message presented to participants may influence how much information these participants disclose. Specifically, we expected that individuals who are shown messages emphasising learning or social aspects of participation will disclose more information in the Sensitive Item part of the survey (than individuals who were shown messages emphasising contribution or altruism).

To conduct this study, we could not tell you in advance that we are studying privacy as that could have made you self-conscious about how you answer questions and eliminated the possibility of studying spontaneous and authentic behaviour. These are important issues and as outlined already, it is important to understand people's behaviour so that we can help create fairer disclosure environments for future generations of citizen scientists, as well as building greater data privacy literacy among the wider population.

We want to assure you, however, that the purpose of these deeply personal queries - this refers to questions from 9.1. 'Did any of your loved ones die while you were growing up?' through to question 9.14. 'May we obtain a copy of your insurance claims (e.g., car, house)?' - was solely to learn WHETHER you decide to respond. This means that after recording whether you gave a response or chose not to answer a particular question, any actual information you gave in response to questions 9.1.-9.14., will be deleted. Such information will not be held nor processed.

Our study is similar to one that has already been published that explored what data people were prepared to disclose in the context of a credit card application. More information on that prior work can be found at [Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card

Application. - UCL Discovery] (http://discovery.ucl.ac.uk/1349987/). The motivational messages were adapted from a previous study on recruitment messages in citizen science and the original text can be found at [Recruiting messages matter: Message strategies to attract citizen scientists | Kevin Crowston] (https://crowston.syr.edu/node/669).

We are thankful for your participation. Your survey responses will be stored securely and separately from any personal identifiers you may have disclosed, such as email address, and we will ensure that all the requirements of the Data Protection Act 2018 and the General Data Protection Regulation are complied with.

If there is any reason why you have changed your mind about being part of this project, please contact us at {email address}, and we will remove your data immediately. You do not have to state a reason for withdrawing from the study. If you have any other questions, thoughts or comments please do not hesitate to contact us. We welcome and are grateful for all feedback.

The lead supervisor for this study is Professor Anna L. Cox, UCLIC, University College London, 66 - 72 Gower Street, London, WC1E 6EA; telephone: {phone number}; email: {email address}.

A.7.5. Debriefing message Study 5

<u>EXPERIMENT DEBRIEF INFORMATION</u>

‘INFORMATION DISCLOSURE IN CITIZEN SCIENCE’

Dear Participant,

Thank you for completing this survey. We would like to give you more information about the purpose and design of this study. This survey served two goals. Firstly, it helped us gather information about how individuals answer questions about their experience of pet ownership. Secondly, we wanted to learn about how different individuals respond to questions about very personal information concerning themselves and their friends or family.

We aim to clarify whether the motivations that drive people to participate in citizen science can also influence how much personal information people will share while taking part in citizen science projects. Such research is of particular importance following the recent disclosure of ‘breach of trust’ and potential misuse of Facebook data. If we are to help design safeguards that protect individuals from their data being compromised, we must first understand when, how, and why individuals make decisions to share personal data. We hope that this work will help researchers to design citizen science projects in future in a way that is more aligned with people’s attitudes and preferences in relation to data privacy and data disclosure.

After consenting to take part in the survey, each participant was asked to answer a question about which motivation to participate in citizen science is most important to them: 1. Learning about science, 2. Social interaction, 3. Making a contribution to a project, or 4. Helping scientists. Then, participants were asked to answer 4 Demographics-related questions (enquiring about age, gender, use of Internet and previous citizen science participation), as well as 14 Neutral Items and 14 Sensitive Items. The Neutral Items were questions about the experience of having a pet. The Sensitive Items were questions about different sources of life, relationship and financial stress.

In this study we set out to test two hypotheses. The first hypothesis assumed that people will be more keen to share information when answering neutral questions than when answering sensitive questions. Secondly, we hypothesised that the degree to which people disclose data when answering sensitive questions will vary depending on what motivated them to take part in citizen science in the first place - for example, based on previous research, we expect that individuals motivated by Learning about science will disclose more sensitive information than those motivated by Social Interaction, Making a contribution to a project, or by Helping Scientists.

To conduct this study, we could not tell you in advance that we are studying privacy as that could have made you self-conscious about how you answer questions and eliminated the possibility of studying spontaneous and authentic behaviour. These are important issues and as outlined already, it is important to understand people’s behaviour so that we can help create fairer disclosure environments for future generations of citizen scientists, as well as building greater data privacy literacy among the wider population.

We want to assure you, however, that the purpose of these deeply personal queries - this refers to questions from 8.1. ‘Did any of your loved ones die while you were growing up?’ through to question 8.14. ‘May we obtain a copy of your insurance claims (e.g., car, house)?’ - was solely to learn WHETHER you decide to respond. This means that after recording whether you gave a response or chose not to answer a particular question, any actual information you gave in response to questions 8.1.-8.14., will be deleted. Such information will not be held nor processed.

Our study is similar to one that has already been published that explored what data people were prepared to disclose in the context of a credit card application. More information on that prior work can be found at [Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card Application. - UCL Discovery] (http://discovery.ucl.ac.uk/1349987/). The answers to question 5.1. ('Which benefit of taking part in citizen science is most important to you?'), were adapted from a previous study on recruitment messages in citizen science and the original text can be found at [Recruiting messages matter: Message strategies to attract citizen scientists | Kevin Crowston] (https://crowston.syr.edu/node/669).

We are thankful for your participation. Your survey responses will be stored securely and separately from any personal identifiers you may have disclosed, such as email address, and we will ensure that all the requirements of the Data Protection Act 2018 and the General Data Protection Regulation are complied with.

If there is any reason why you have changed your mind about being part of this project, please contact us at {email address}, and we will remove your data immediately. You do not have to state a reason for withdrawing from the study. If you have any other questions, thoughts or comments please do not hesitate to contact us. We welcome and are grateful for all feedback.

The lead supervisor for this study is Professor Anna L. Cox, UCLIC, University College London, 66 - 72 Gower Street, London, WC1E 6EA; telephone: {phone number}; email: {email address}.