

Empirically Studying Research Ethics with Interface Designs for Debriefing Online Field Experiments

Jonathan Zong

Princeton University
Department of Computer Science

May 2018

Advised by
J. Nathan Matias (Department of Psychology)
Marshini Chetty (Department of Computer Science)

This thesis represents my own work in accordance
with University regulations.

Abstract

Debriefing is an essential research ethics procedure in non-consented research wherein participants are informed about their participation in research and provided with controls over their data privacy. This paper presents a novel system for conducting and studying debriefing in large-scale behavioral experiments on online platforms. I designed a debriefing system, with an accompanying evaluation study, which are both delivered as a web application. I recruited 1182 users on Twitter who have been affected by DMCA takedown notices into an empirical study on debriefing. The key contributions of this paper are 1) the design and implementation of the debriefing system, 2) empirical findings from the debriefing study on its unexpectedly low response rate, and 3) an evidence-based analysis of challenges researchers face in recruiting participants for research ethics and data privacy research.

Acknowledgements

I am first and foremost, always, immensely grateful for the support of my family—my parents Helen and Shuh, and my sister Janet. They are unconditionally on my team no matter what I choose to pursue, generous listeners, role models, sources of wisdom, thoughtful conversation partners. I could not ask for more.

I am incredibly fortunate to have had the opportunity to work with and learn from Nathan Matias this past year. Nathan is a mentor in the truest sense of the word—extremely generous with his time, attention, and experience, with a canny understanding of and empathy for the kind of guidance I needed to grow as a researcher this year. The influence of his intellectual work and support is reflected in each of these pages.

Through working with Nathan, I've been fortunate to engage with wider communities in research, like CivilServant and the Paluck Lab. I'm also grateful to have met Merry Mou and Jon Penney, whose work on the DMCA study provides an important setting for this debriefing work, through this project.

Thank you to the friends who have supported me with laughter, conversation, shared presence, and shenanigans. The people I've become close with at Princeton are what makes it feel hardest to leave. This includes the overlapping group chats of Bee Hell in Princeton Feminist Club, tbh same, lunch club, and many other individuals who are too numerous to name.

Finally, I'm grateful for other mentors I've had, whose presence in my intellectual development is no small contribution to this work: David Reinfurt, Jane Cox, Jeff Snyder, Aatish Bhatia, Tal Achituv, Judith Hamera, Don Adams.

Abstract	2
Acknowledgements	3
Introduction	7
Debriefing and the user experience of research ethics procedures	7
Empirical research on research ethics	8
Goals of this project	9
Research Ethics	10
What people need from research ethics procedures	10
The user experience of debriefing	11
Figure 2.1. User flows through research with and without consent [12]	12
Debriefing and research design	13
Deception-based research	14
Non-consented research	14
Evaluating the ethics of research procedures	16
Opt out rate	16
Risks and benefits from the intervention	17
Privacy	18
Design Considerations for Debriefing Systems	18
Informing users	18
Providing users the ability to opt out	19
Framing and defaults	21
An Interface for Debriefing Experiments	21
Features of the system	21
Debriefing interface	22
Figure 4.1. Debrief interface: table of data collected in the study	23
Figure 4.2. Debrief interface: visualization of study results	23
Figure 4.3. Debrief interface: opt out controls	24
Technical details	25
Evaluation Study	25
The DMCA context	25
Goals of the debriefing evaluation study	25
Recruitment methods and goals	26
Survey interface	27
Figure 5.1. Survey interface: consent page	27

Figure 5.2. Survey interface: Twitter authorization	28
Figure 5.3a. Survey interface: randomization (control)	30
Figure 5.3b. Survey interface: randomization (treatment)	30
Figure 5.4. Survey interface: survey questions	31
Survey infrastructure	31
Figure 5.5. Survey interface: automated compensation	32
The Recruitment Problem	32
Recruitment procedure	32
Table 6.1. First recruitment attempt	33
Table 6.2. Recruitment message variations	35
Table 6.3. Recruitment attempts during main study period	36
Table 6.4. Recruitment and participation	37
Recruitment response	37
Figure 6.1. Consent page of the survey web application	38
Some hypotheses for low response rate	39
Future work with debriefing	42
Different models of consent	42
Forecasting	43
Conclusion	44
Bibliography	47
Appendix	49
Code for the debriefing system	49
Forecasting and debriefing study pre-analysis plan	49

1. Introduction

As behavioral experimentation becomes more widespread in society through online platforms, we need new ways to manage the ethics and accountability of that research. Since this research is delivered digitally, we can develop novel technologies for managing large-scale research ethics. Because models of consent and accountability in research ethics involve communicating complex ideas to the public, advances in user interfaces for managing participation in research can contribute to novel approaches in research ethics.

For example, in large-scale academic experiments online, due to practical concerns obtaining informed consent from the entire population is not always possible. Under the Common Rule, a university IRB can waive the requirement for a signed consent form by the following criteria: the study must have minimal risk, obtaining informed consent must be impractical, and there must be a post-experiment debriefing [6].

1.1. Debriefing and the user experience of research ethics procedures

Debriefing is a procedure in experiments involving human subjects wherein, after the experiment has concluded, participants are provided with information about the experiment and the data that was collected in the process. The procedure serves an important ethical purpose by giving the participants an opportunity to clarify their involvement, ask questions, or opt out; this is especially important in experiments where there was any form of deception or where informed consent was not obtained beforehand.

Research ethics procedures like debriefing can be understood from an HCI standpoint as an essential part of the user experience of being included in a study. Because successful debriefing requires people to understand the experiment and in some cases make important decisions, novel user interface approaches may improve the debriefing process.

1.2. Empirical research on research ethics

A field experiment is a study which makes interventions and observations in the world, as opposed to in a lab or with a survey. An experiment from 2012 in which “Facebook showed some users fewer of their friends’ posts containing emotional language, then analyzed the users’ own posts to see whether their emotional language changed” is an example of a field experiment that has an intervention (hiding posts), has an observation (analyzing users’ own posts), and is situated in everyday life (normal usage of Facebook) [6]. It is also an example of a field experiment that prompted outcry from the public due to a lack of ethics and accountability procedures.

In recent proposed standards for the ethical design of field experiments, Desposato recommends debriefing as one standard that serves as a constraint to hold researchers accountable by asking them to consider possible participant reactions up front when designing experiments [3]. In related work examining large-scale experiments on social media users, Grimmelmann gives examples of instances in which debriefs were not included in experiments by corporate entities, resulting in conflicts of interest and lack of transparency from participants’ points of view [6].

Desposato argues that “as a discipline we should engage [research ethics] issues directly and work toward shared norms” in order to ensure subjects’ protection and the viability of field experiments as a method [3]. The approach of conducting empirical research on how people make sense of different kinds of research ethics procedures plays an important role in contributing evidence-based arguments to this conversation. For example, Desposato has done empirical work surveying researchers and participants on the use of informed consent [4]. The evidence he has gathered suggests possible ways to proceed responsibly with non-consented research, making progress on seemingly intractable ethical issues around the increasingly widespread use of large-scale field experiments without informed consent.

1.3. Goals of this project

The goal of this project is to make progress on research ethics by 1) outlining some goals and considerations for the design of user interfaces for debriefing that help people understand what it means to have participated in large-scale online behavioral research, 2) implementing a debriefing system with these considerations in mind, and 3) evaluating the interface with an empirical study. The interface presented in this project instantiates principles of research ethics based on *consent* and *accountability*. These two ideas guide the design discussion in subsequent sections.

In the course of this project, I implemented a debriefing user interface as a web application that can be delivered as a URL to participants in non-consented research. I designed and ran an evaluation study consisting of a survey asking randomly sampled representatives of a group to give feedback the interface and report what their responses might be in a hypothetical

debriefing scenario. The surveys accompanying the debriefing interface are implemented as part of the same web application, which is hosted on *cs.princeton.edu* servers. I wrote code for supporting study systems like automated study recruitment on Twitter and automated compensation on completion of the survey using Paypal, and ran recruitment for about 3 months. Few people responded to the recruitment materials in the study, and in this report I explore the challenges and considerations of recruiting participants for research ethics and data privacy research.

2. Research Ethics

2.1. What people need from research ethics procedures

All debriefing procedures work to achieve at least two key outcomes, summarizing “norms of informed consent and respect for subjects’ autonomy” [4]:

- 1) *Informing*. Researchers want to ensure a state of understanding in participants so that they have comprehension of what is at stake in the research: the questions, data, and any risks or benefits they might incur. This is a task normally fulfilled by *informed consent*, which is part of why debriefing is essential in research that does not use consent.
- 2) *Providing controls*. Once they are equipped with the information and understanding to reason about their personal risks and benefits from engaging in research, we give them control to exercise their right to withdraw from the experiment by opting out. This is an

important mechanism of *accountability* to ensure researchers maintain respect for the participants' autonomy.

Since keeping participants uninformed and out of control undermines public trust, debriefing also contributes to the maintenance of public trust in research. As Desposato notes, “there are practical consequences to ignoring subjects' preferences. One is that we jeopardize public trust in the research enterprise, which may deter participation in all types of studies” [4].

In this project, I focus on debriefing for studies with minimal risk. The federal regulations on human subjects research known as the Common Rule define minimal risk as when “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests” [14]. This definition is the standard that IRB uses to determine whether risk is low enough to waive the informed consent requirement. With regard to controls over withdrawing from the study, this project focuses on opting out as it relates to data sharing and privacy concerns.

2.2. The user experience of debriefing

In a typical field experiment using informed consent, the user flow begins with informed consent and leads into research participation and the conclusion of the research; however, in research that does not use consent, users begin already participating in the experiment without their knowledge (Figure 2.1). Debriefing happens to participants who have not yet gained awareness or understanding about the research. The experiment intervention and data collection happens

before they are debriefed; therefore, the role of informing participants and giving them a choice normally fulfilled by the informed consent step must happen after the experiment instead of before.

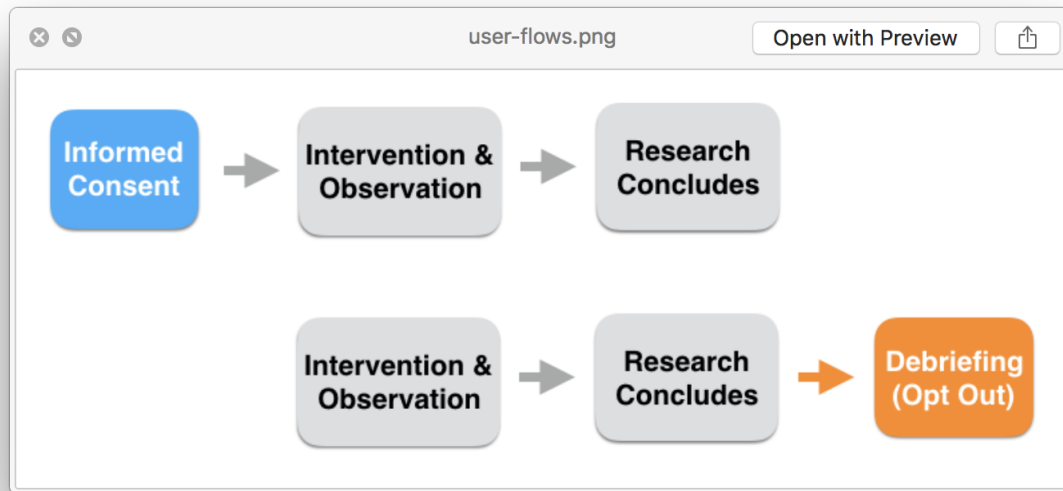


Figure 2.1. User flows through research with and without consent [12]

A debriefing usually communicates the following points: what was being studied, how participants were deceived, why the non-consent was necessary, the study methods and results, the procedure for opting out of the study, and any further resources useful for the participant.

In lab experiments, debriefing typically happens either with a debriefing form or a conversation script for in-person debriefing. Lab debriefing has advantages due to the participants' in-person access to a researcher, enabling the possibility for the researcher to "assess a subject's state and therefore to determine whether an individual has been upset by an experimental procedure or understands feedback received" [9]. This is very useful for

researchers to understand if they have succeeded in informing the user and providing the care for which they are accountable.

In online research, debriefing is usually delivered over web interface or email. Web interfaces allow debriefing to be more individually tailored to each participant. Most literature on online debriefing is in the context of online survey research—participants are on a web page specifically designed for an experiment, and are debriefed at the end of the web page’s user flow. Because participants are able to drop off from the study before reaching the end, researchers face unique challenges in ensuring their users are requisitely engaged in order to inform them. Some researchers in the past have created interface solutions to maximize the likelihood that the debriefing information will be received. For example, “researchers can deliver debriefing material through a link to a ‘leave the study’ button or through a pop-up window, which executes when a subject leaves a defined Web” [9].

My system differs from online survey debriefing systems because it debriefs research that does not necessarily occur on a study web page, but instead is a field experiment situated on a social media platform, coinciding with everyday use.

2.3. Debriefing and research design

What are the features of a study’s design that make the use of debriefing necessary? Debriefing is most commonly used in deception-based research; however, while deception is a broad category, it does not encompass all research that does not use consent. For that reason, I use the term *non-consented research* to make a more precise claim about the scope of research designs considered in this project.

2.3.1. Deception-based research

Studies that waive informed consent usually include some element of deception—a word that can mean different things in different fields. Deception can mean actively lying to participants, or it can mean withholding information or incomplete disclosure.

For example, in economics Cooper explains that deception is “generally taken to encompass instructions or materials that actively mislead subjects by stating or strongly implying something that is not true.” Economists have a “de facto ban on the use of deception,” but because this understanding of deception denotes active lying but not omission or ambiguity, “other experimental techniques that could arguably be classified as deception are considered acceptable” [2]. In the economics context, active deception is considered detrimental because it “potentially undermines the experimenter’s control over [subjects’ economic incentives]” and therefore threatens the validity of the experiment results [2]. There is a consistent consensus that actively providing false information constitutes deception, and “this consensus is also shared across disciplinary borders” including in psychology [7].

2.3.2. Non-consented research

The study used in the design of this debriefing system does not involve lying to participants, although the debriefing system could conceivably be used in that situation as well. Here the debriefing system is applicable to situations where researchers are not notifying people that they are part of a study. To make this distinction more precise and to differentiate it from other forms of research which might use debriefing, I introduce the term *non-consented research*. In

non-consented research, participants are not actively misled, but are kept unaware that they are included in a field experiment even though this is contrary to their default expectations.

Why is a new term necessary? Most literature about deception talks about active deception, but the situation in large-scale online research is more commonly an omission to disclose and consent. While this is closer to discussions of the ethics of withholding information or providing incomplete disclosure, it does not exactly match that context either. Most literature on withholding information focuses on situations where, for example, researchers do not “[acquaint] participants in advance with all aspects of the research being conducted, such as the hypotheses explored and the full range of experimental conditions” [7]. Most of these discussions start from a baseline understanding between the participant and the researcher that research is happening, because these discussions most commonly contend with lab or survey experiments rather than online field experiments.

Psychologists Hertwig and Ortmann propose an alternative notion of deception defined as a violation of participant expectations:

Although deception is commonly defined on the basis of the experimenter’s behavior (e.g., intentionally providing false information), one could define it alternatively on the basis of how participants perceive the experimenter’s behavior. According to such a definition, deception would have occurred if participants, after being completely debriefed, had perceived themselves as being misled. Such an approach defines deception empirically and post hoc rather than on the basis of norms devoid of context. [7]

This empirical notion of deception based on participants’ default assumptions helps motivate what I mean by non-consented research—for example, large-scale online research may be violating the assumption of not being part of a field experiment while using Twitter. It is important for our thinking about the ethics of both deception and non-consent to be centered on participant expectations and state of understanding, as a way for researchers to be held accountable to participants.

2.4. Evaluating the ethics of research procedures

Recent work in research ethics has suggested some paths toward evaluating the ethics of research procedures. For example, Desposato has surveyed researchers and research participants to gather data about how each group thinks about the ethics of political science experiments that waive consent. This is an example of what he calls the “empirical ethics” approach to understanding what participants think about research, which helps researchers learn both the acceptability of their research and the potential consequences of violating the public’s expectations [4].

In evaluating a debriefing interface, what measures should be used to understand its effectiveness? The following are some measures that this project considers. More detail about the specific methodology of the study is given in Section 5.

2.4.1. Opt out rate

Opt out rate is an important measure both because it is important to researchers in terms of the usefulness of the data they collect, and also because it is an important signal from participants to

researchers about their state of understanding about possible risks and benefits of the research. It also measures the ability and willingness of participants to exercise their right to withdraw from research. One might hypothesize that if a debriefing interface is successful at informing users and increasing their understanding of the research and data collection, it would have some effect on the opt out rate.

But is a lower opt out rate always a good thing? There are drawbacks to evaluating a system solely through opt out rates. For instance, looking only at opt out rates does not include context on the relative risks and benefits of the experiment—concerns which should normatively affect the opt out rate. Consider a study where participants are part of a vulnerable population and sensitive data is collected which would be harmful if made public. In this instance, the most desirable opt out rate is probably lower than a study with minimal risk. But imagine a medical study with equally sensitive data, but with a high potential to directly benefit the well-being of the participants—maybe the desired opt out rate is higher in this instance. Opt out rates must be contextualized by expectations of how high or low we desire them to be, based on considered balancing of risks and benefits. It is important not to think of them as a standalone metric to be minimized or maximized.

2.4.2. Risks and benefits from the intervention

Participants' understanding of the risks and benefits of interventions in research is also reflective of how effectively a potential debriefing interface has informed them. To measure this, we can ask participants for their assessment of whether answering the questions posed in the study would benefit society, and whether the answers would benefit them personally. To get a sense of

their risk perceptions, we can ask participants whether they feel positively or negatively about being included in the study.

2.4.3. Privacy

It is also important to have a measure of users' attitudes about the privacy of their data. One useful measure to help understand this is to ask participants about their surprise at the data collection that happened in the research. Are participants aware that such data collection is possible, and to what extent do they understand it to be commonplace? Prior work by Fiesler and Proferes has shown that most users are not aware that public tweets could be used by researchers, nor do they feel positively about this use [5]. These participant expectations, and whether any data collection stays within the horizon of their expectation, are central to understanding whether participants feel that their privacy has been violated.

3. Design Considerations for Debriefing Systems

Even though debriefing systems work to increase participant autonomy, researchers still make decisions that affect the range of possible responses and actions for participants. I will discuss some general considerations that anyone designing a debriefing must contend with, which have become apparent through the process of designing the system presented in this project.

3.1. Informing users

Any debriefing system will need to communicate the details of an experiment. It will also need to communicate how the experiment affected a participant personally, through intervention or

through data collection. For different pieces of information, different approaches might be more effective. When presenting data, researchers might consider the following choices:

- 1) *Text-based and/or visual.* Different types of information are clearest as text, or in a table, or as an image, or even a combination of these. In this study, I include the presence or absence of tables and visualizations as condition variables in the evaluation to see if they have a measurable effect on user understanding.
- 2) *Personal and/or collective.* Is it most straightforward to only show participants their own data and nothing more, or might showing analysis about how they stand in relation to others in the study prompt them to contextualize their participation as a contribution to a collective research question? In this study, I choose to show two graphs, one showing the effect on the participants' tweets per day and one showing the effect on all participants' tweets per day, on average.

These decisions have to do with way information is delivered, which is inextricable from participants' ability to understand it. It is important to consider—and possibly even empirically test—what approaches are most helpful toward the goal of informing users and advancing their understanding of the research.

3.2. Providing users the ability to opt out

When asking users whether they intend to opt out, it is critical to decide what exactly we are asking them to opt to. Including users in online research that necessitates debriefing usually

involves potentially making an intervention in their online experience without their knowledge, and collecting data on them before and after to measure a possible effect. At the point in time when they receive debriefing, any intervention would already have been made; therefore, we are asking them to make a decision on how we treat the data that we have collected. When users choose to opt out, what are the possibilities that allow researchers to satisfy participants' intent while still allowing them to answer potentially beneficial research questions? Within this tradeoff, there are different possible scopes to the action of opting out. When researchers honor an opt out request, it could mean entirely deleting the participant's data, but it could also mean opting out of data sharing with other researchers, opting out of public data sharing, or opting in to anonymization and obfuscation. All of these options have different potential consequences for participant privacy and for the goals of the research.

What are the different implications of these choices for research ethics? There are advantages to sharing datasets between researchers that many in the open science community advocate for. For example, sharing datasets allows for transparency into the research process that others can, for example, learn from or audit for accountability purposes. It also allows future researchers to reproduce experiments to either further validate or bring into question the results of the original experiment. But as Ed Freeland points out, "publishing data introduces privacy risks for participants in research. While US legislation HIPAA covers medical data, there aren't authoritative norms or guidelines around sharing that data" [11].

How anonymous can research participants reasonably expect their data to be if opting out causes their data to be anonymized instead of fully deleted? Ed Freeland notes that "the landscape of data re-identification is changing from year to year, but the consensus is that

anonymization doesn't tend to work" [11]. Therefore, from a research ethics standpoint we should assume that the choice between full deletion and anonymization is a choice of different risk tolerances for participants.

3.2.1. Framing and defaults

The way the decision about personal data is presented to participants will influence how they respond. In other words, is the decision to withdraw presented as opt out or opt in? There is a well-documented "tendency of decision-makers to view the default as the standard of comparison, or as the popularly endorsed, or correct answer," and as designers we must be aware that "the form of the question produce[s] sizable differences in participation" [1].

When designing a debriefing interface, choosing a no-action default is unavoidable. For instance, not everyone will click on the link to open the debriefing. How do we handle these participants' personal data with care despite our lack of feedback from them?

If the default is to retain the data, then we might expect a higher retention rate; conversely, if the default is to remove the participant, then we might expect a higher withdrawal rate. Which of these is more desirable depends on the context of the research, and its relative risks and benefits.

4. An Interface for Debriefing Experiments

4.1. Features of the system

The debriefing system proposed by this project is a web application that was deployed to *dmca.cs.princeton.edu* for the duration of this study. It has three main parts: 1) the debriefing

interface, 2) an evaluation survey interface, and 3) survey infrastructure including scripts for automated recruitment and compensation.

4.1.1. Debriefing interface

The goal of the debriefing interface is to inform users and give them control over their data privacy. In addition to text explanations, two main features support the goal of informing users about their participation in the study. The first feature is a table in the debriefing interface which displays all of the data collected on the participant (Figure 4.1). The intent is to be transparent and precise about data collection so that the participant can decide whether the data is within acceptable bounds of their privacy expectations, interesting, or potentially useful. The second feature is a visualization illustrating some results from the study (Figure 4.2). In addition to what the participant also needs to know why the data was collected. Contextualizing their data as a contribution to the results of the overall study helps communicate the potential relevance and value of the results to them personally and to society in general.

Here is what we collected about your public Twitter behavior from [date] to [date]. This information helps us find out the effect of providing legal information on average. We will not share any of your data, and our dataset includes only the following information and no more:

Do you use the default Twitter profile image or your own?	no
How many lifetime tweets had you sent before receiving the copyright notice?	1007
Do you have a <input checked="" type="checkbox"/> verified account ?	no
How many years has your account existed?	6
What language is your Twitter account associated with?	en
Was the link we sent you clicked on ?	(filled in from data)
Was your account suspended at all during the study period?	(filled in from data)
Was your account deleted at all during the study period?	(filled in from data)
Was your account protected at all during the study period?	(filled in from data)
How many copyright notices did you receive during the study period?	(filled in from data)
How many tweets per day did you post before receiving your first copyright notice?	(filled in from data)
How many tweets with pictures or other media did you post per day before receiving your first copyright notice?	(filled in from data)
How many tweets per day did you post after receiving your first copyright notice?	(filled in from data)
How many tweets with pictures or other media did you post per day after receiving your first copyright notice?	(filled in from data)

To learn more about our research, you can read our study design online [link].

Figure 4.1. Debrief interface: table of data collected in the study

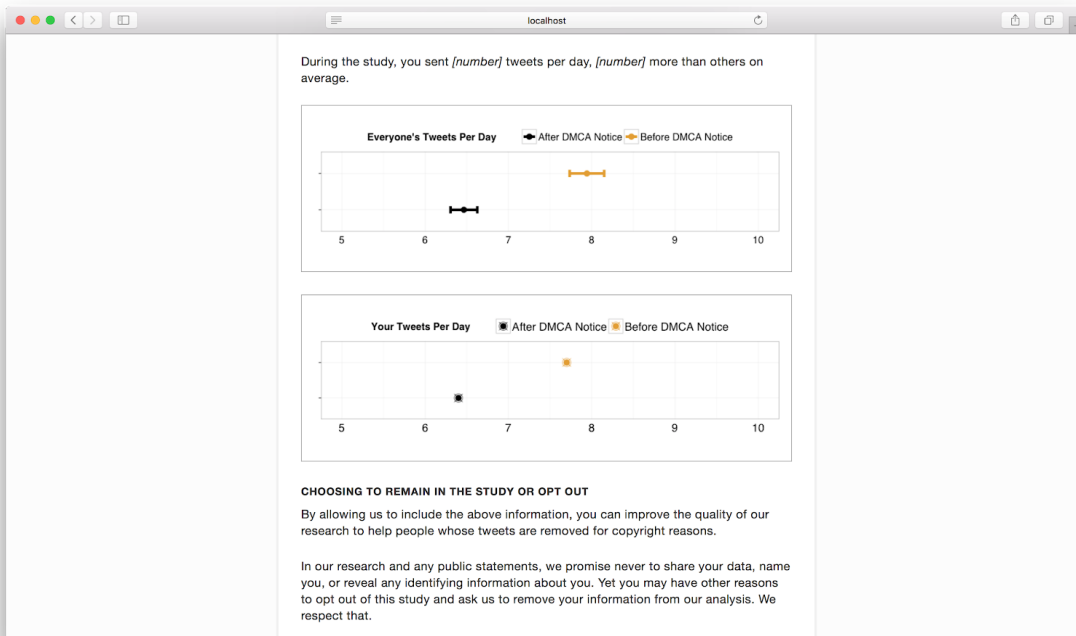
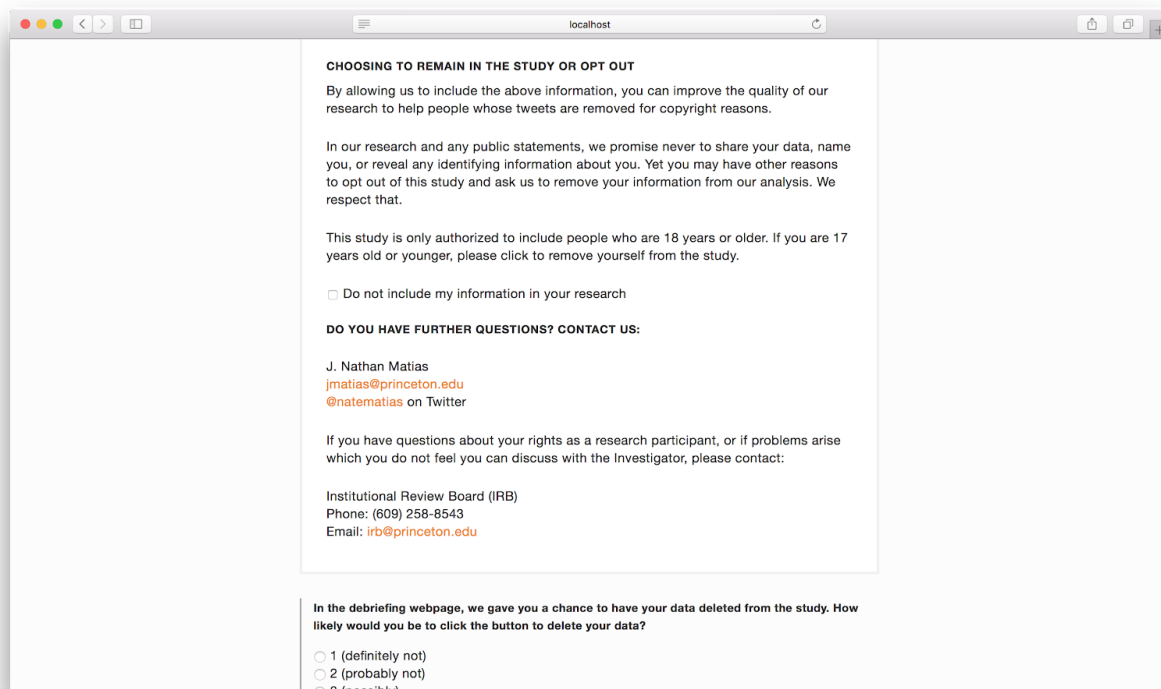


Figure 4.2. Debrief interface: visualization of study results

The main feature in the debriefing interface that supports the second goal of providing users control over their own participation is an opt out checkbox (Figure 4.3). Because the decision to opt out is presented below the parts of the interface designed to inform, ideally the participant will possess the understanding of their relationship to the research to make accurate assessments of their own potential risks and benefits. The better they understand these factors in their decision-making, the more successful we as researchers have been at fulfilling our ethical obligations to them.



The screenshot shows a web browser window with the address bar set to 'localhost'. The page content is as follows:

CHOOSING TO REMAIN IN THE STUDY OR OPT OUT

By allowing us to include the above information, you can improve the quality of our research to help people whose tweets are removed for copyright reasons.

In our research and any public statements, we promise never to share your data, name you, or reveal any identifying information about you. Yet you may have other reasons to opt out of this study and ask us to remove your information from our analysis. We respect that.

This study is only authorized to include people who are 18 years or older. If you are 17 years old or younger, please click to remove yourself from the study.

☐ Do not include my information in your research

DO YOU HAVE FURTHER QUESTIONS? CONTACT US:

J. Nathan Matias
jmatias@princeton.edu
[@natematias](#) on Twitter

If you have questions about your rights as a research participant, or if problems arise which you do not feel you can discuss with the Investigator, please contact:

Institutional Review Board (IRB)
Phone: (609) 258-8543
Email: irb@princeton.edu

In the debriefing webpage, we gave you a chance to have your data deleted from the study. How likely would you be to click the button to delete your data?

☐ 1 (definitely not)
☐ 2 (probably not)
☐ 3 (possibly)

Figure 4.3. Debrief interface: opt out controls

4.2. Technical details

The debriefing system is implemented in Python 3. It uses the Flask web application framework with the SQLAlchemy database toolkit and Alembic database migration framework. Further details on the code can be found in Appendix A.

5. Evaluation Study

5.1. The DMCA context

The evaluation of the debriefing interface will take place in the context of a research project empirically studying the effects of copyright enforcement on Twitter, conducted by Jon Penney and Merry Mou. The debriefing interface will be used to debrief participants who were included in this research about automated copyright enforcement. This means that all users who will be giving feedback on the debriefing are part of the population of users who have received DMCA takedown notices on the Twitter platform within 2 months prior to the beginning of the debrief evaluation study.

5.2. Goals of the debriefing evaluation study

In this study, we ask Twitter users who have received DMCA copyright notices in the past to give feedback on a web interface for debriefing participants in field experiments. We also survey them about research ethics and their choice to opt out of the research. The full pre-analysis plan for this study can be found in Appendix B.

5.3. Recruitment methods and goals

In this study, I recruit Twitter accounts that have received copyright notices in the two months prior to the beginning of the pilot. These notices are a matter of public record in the Lumen database, a service operated by Harvard University researchers that is independent of Twitter, and are available as a web service at <https://lumendatabase.org/>.

Participants will be included in this study if they appear in the Lumen database of Twitter DMCA takedown notice, if the database records that they received a DMCA takedown notice within the past two months, if we identify links in the notice to this participant's Twitter account, if we can successfully identify that Twitter account via the Twitter API, and if Twitter reports that the account has a "en" language (which is a proxy for locale).

Participants are recruited by @-messages sent to their twitter account, with a link to the debriefing interface and survey. Participants are then compensated for completing the survey.

The survey asks participants to imagine that they had been part of a field experiment. It shows participants a debriefing interface, grouping participants into a stratified sample of people whose content was removed for copyright reasons, and those whose content was permitted to remain on Twitter, to ensure balance across experiment arms between both groups. The system randomly assigns participants to variations: 1) whether they are assigned to the control group of the imagined experiment or not, 2) whether the debriefing interface includes a graphic of the results, and 3) whether the debriefing interface includes a table of the collected data.

Throughout the debriefing experience, participants are surveyed to obtain outcome variables, including information about their past experiences, their expected behaviors, and their views on the risks and benefits of the research.

5.4. Survey interface

The web application that delivers the debriefing interface also includes the survey interface needed to run the study that evaluates the debriefing interface. When participants first arrive on the web application, they will see the study consent page, shown in Figure 5.1. This page describes the evaluation study and asks the user to consent to taking the survey by clicking the Twitter login button.

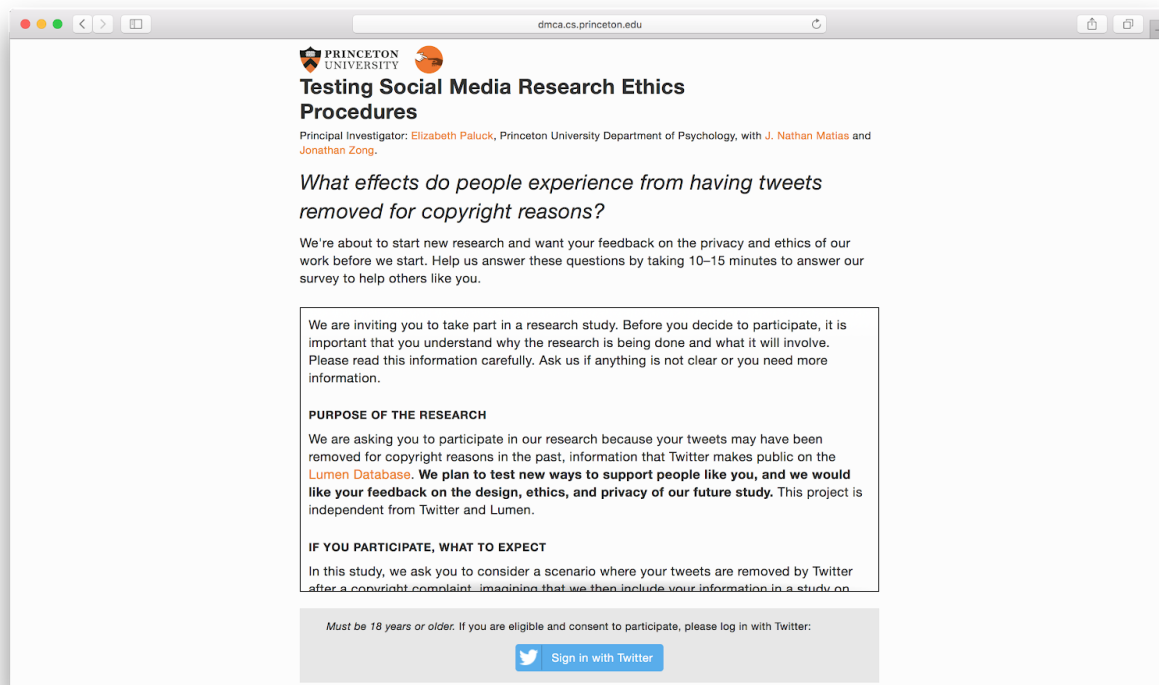


Figure 5.1. Survey interface: consent page

This button redirects to Twitter's authentication page, which requests read-only permissions from the user (Figure 5.2). This is the minimal permission Twitter allows us to request, and lets us observe only publicly available details about the account.

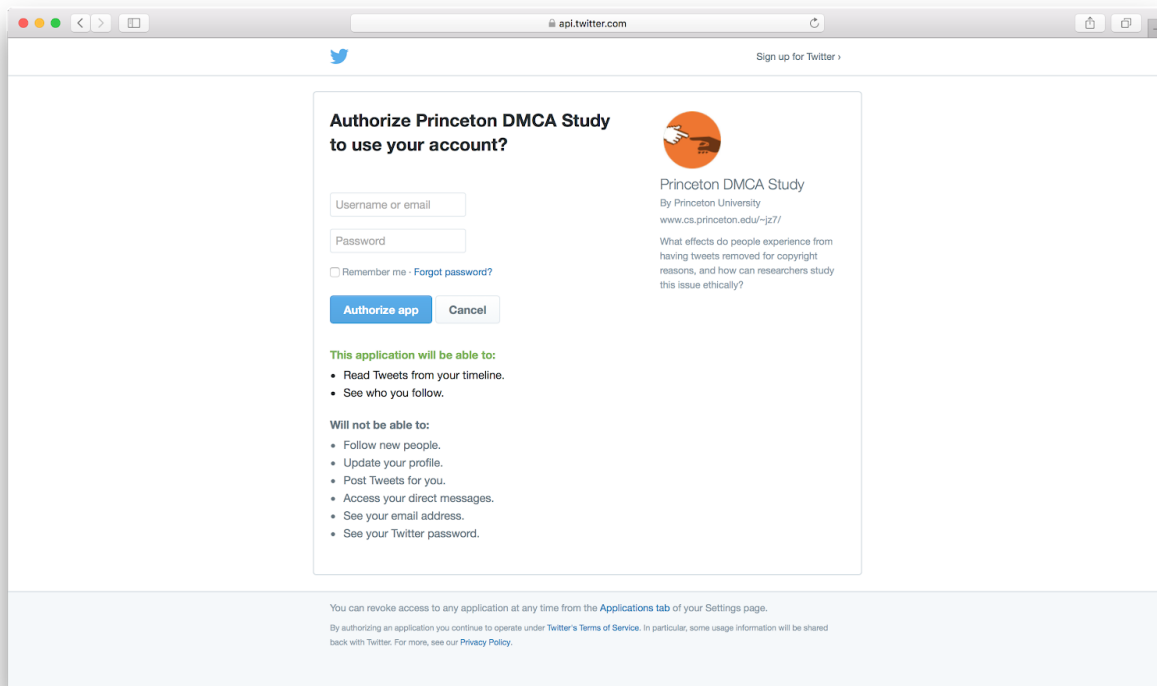


Figure 5.2. Survey interface: Twitter authorization

Once the user has authenticated with Twitter, they will be redirected into the survey. The survey software assigns each user to a randomization once they authenticate. This means that within the survey, users with different assignments can be shown variations of the survey interface. Figure 5.3a. and 5.3b. show an example of a survey question with different prompts

depending on a condition that represents whether the user is shown the control or treatment tweet in a hypothetical debriefing situation.

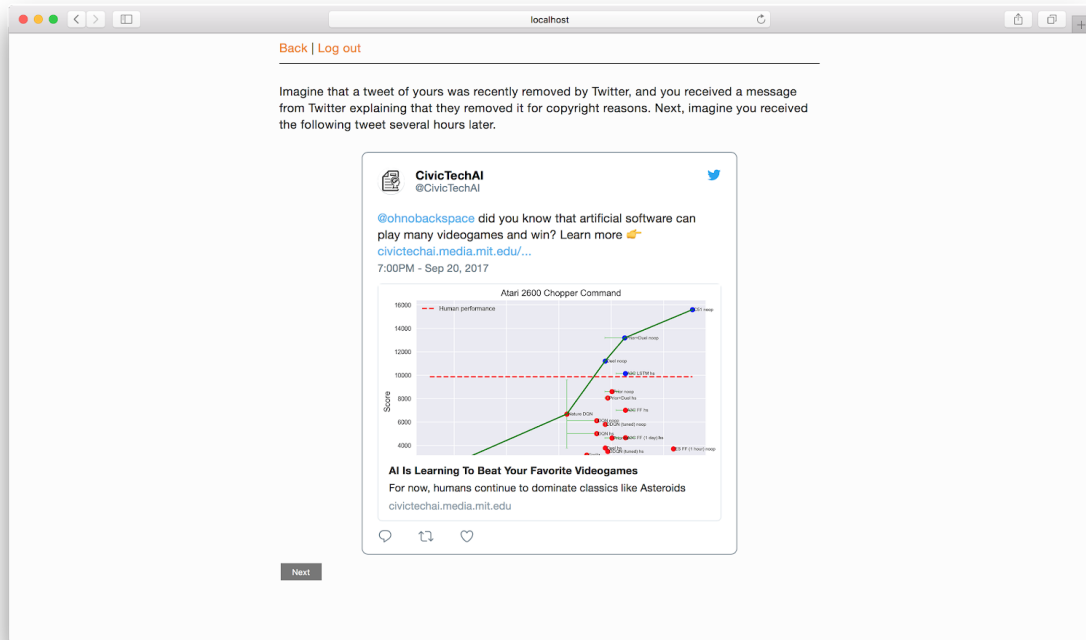


Figure 5.3a. Survey interface: randomization (control)

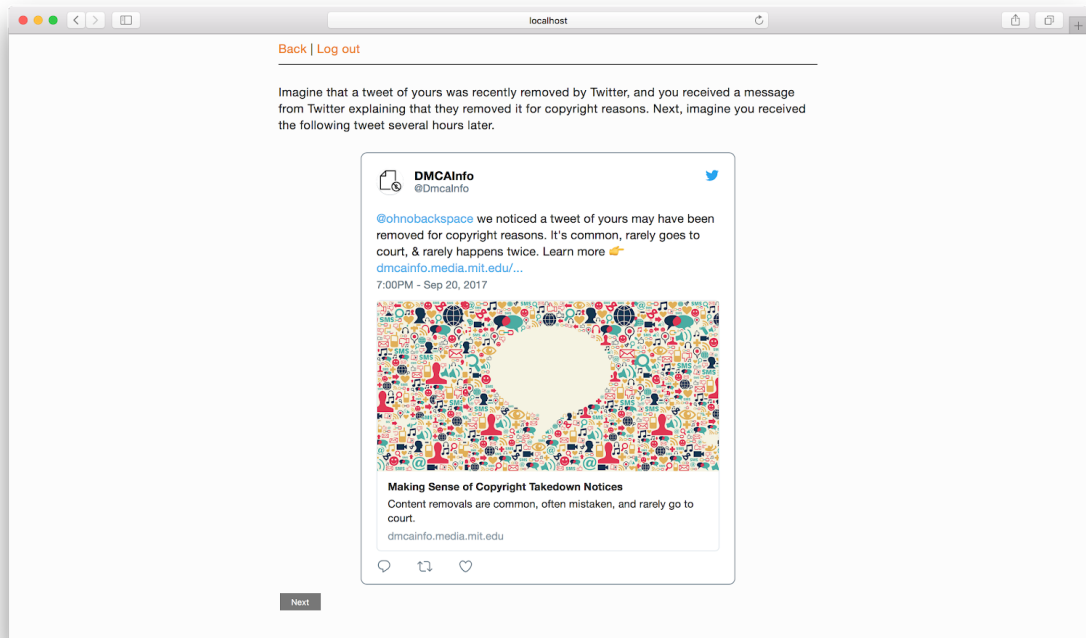
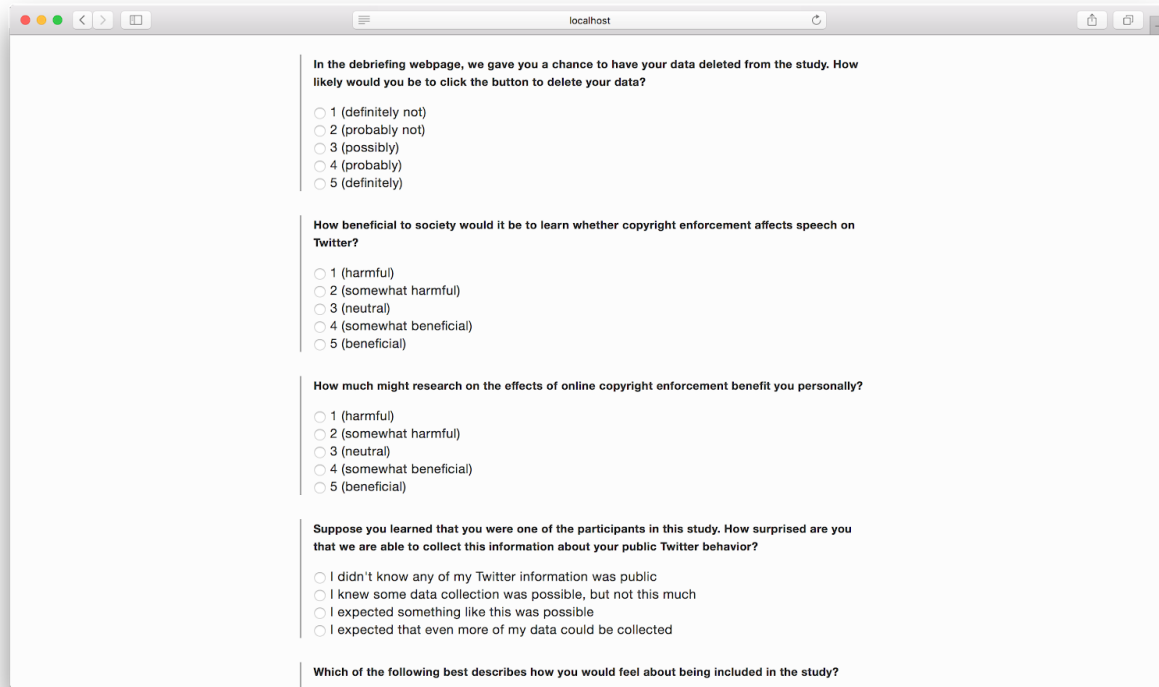


Figure 5.3b. Survey interface: randomization (treatment)

Finally, the survey interface displays survey questions for the participants to answer (Figure 5.4).



The screenshot shows a web browser window with the address bar set to 'localhost'. The survey interface contains five questions, each with a set of radio button options:

- Question 1:** In the debriefing webpage, we gave you a chance to have your data deleted from the study. How likely would you be to click the button to delete your data?
Options: 1 (definitely not), 2 (probably not), 3 (possibly), 4 (probably), 5 (definitely)
- Question 2:** How beneficial to society would it be to learn whether copyright enforcement affects speech on Twitter?
Options: 1 (harmful), 2 (somewhat harmful), 3 (neutral), 4 (somewhat beneficial), 5 (beneficial)
- Question 3:** How much might research on the effects of online copyright enforcement benefit you personally?
Options: 1 (harmful), 2 (somewhat harmful), 3 (neutral), 4 (somewhat beneficial), 5 (beneficial)
- Question 4:** Suppose you learned that you were one of the participants in this study. How surprised are you that we are able to collect this information about your public Twitter behavior?
Options: I didn't know any of my Twitter information was public, I knew some data collection was possible, but not this much, I expected something like this was possible, I expected that even more of my data could be collected
- Question 5:** Which of the following best describes how you would feel about being included in the study?

Figure 5.4. Survey interface: survey questions

5.5. Survey infrastructure

The debriefing software includes additional non-interface features related to the evaluation study. This includes a recruitment script that samples from a list of Twitter user IDs and sends recruitment tweets to eligible accounts. Recruitment is discussed in more detail in Section 6.

The system also includes software for automated compensation. Upon completing the survey, participants are shown a final page with a place to submit their email address. The

system will send a PayPal API request which delivers a link to their email where they can claim their compensation, even if they do not have a pre-existing PayPal account (Figure 5.5).

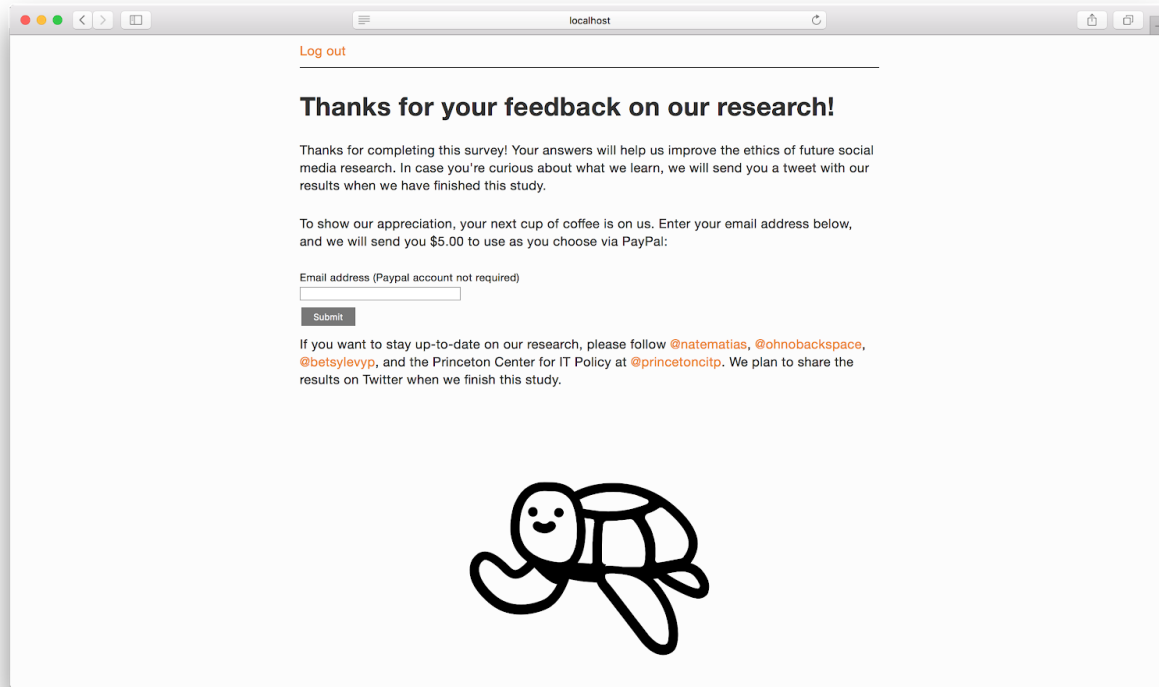


Figure 5.5. Survey interface: automated compensation

6. The Recruitment Problem

6.1. Recruitment procedure

In order for people to be debriefed, they have to actually take up the offer to be debriefed. This turned out to be difficult. Table 6.1 shows the results of the first recruitment attempt for this study between 12/22/17–1/27/18. Out of 399 accounts sampled for recruitment, no one responded. It became clear that I needed to a) be more specific about the sample, b) try a variety

of recruitment options, and c) monitor more information about responses and activity even when people don't complete the survey.

Start– End	Accounts sampled	Compensation	Recruitment message variation	Participated
12/22/17– 1/27/18	399	\$0	A (see Table 6.2.)	0

Table 6.1. First recruitment attempt

With these changes in mind, I restarted the recruitment process by generating a new sample. On 2/25/2018, I ran a script to retrieve records of DMCA takedown notices on Twitter from the prior 60 days. The notices are publicly recorded in the Lumen Database. For the specified time range, the script returned 16763 lumen notices.

After parsing Twitter screen names from these records and deduplicating, the script produced 59034 unique screen names, which I converted to user IDs for easier use with the Twitter API, since screen names sometimes change. I excluded accounts if they were suspended or not found. Of the accounts, 2939 were suspended and 1791 were not found. After excluding these accounts, there remained 54304 valid user entries. The recruitment script began making recruitment attempts from a representative, randomly sampled set of accounts from the list on 3/3/2018.

The recruitment script excluded accounts if they 1) were inactive, which I define as an account that has not sent a tweet in the week prior to when we attempt to tweet at them, 2) if the language of their account was not set to English, 3) if they were suspended, 4) private, or 5) not

found. I included accounts which did not meet any of these exclusion criteria; in this context, including an account in recruitment means sending a tweet at them.

Throughout the process, there were variations made to the recruitment message that the script sent to users. I began the recruitment process with no compensation offer attached to the survey, and gradually tried different compensation rates and messaging to attempt to improve the response rate. Table 6.2 shows different variations on the recruitment message format.

A. No Compensation	B. Compensation
<div data-bbox="219 304 787 346">  DMCAInfo @DmcaInfo Follow </div> <p data-bbox="219 367 787 493">@username Have your tweets ever been taken down for copyright reasons? ©🔥 Help researchers learn how to help people like you</p> <div data-bbox="219 514 787 892">  <p data-bbox="227 808 779 892">DMCA Research - Princeton University We'd like your input on a study to help people who have their tweets removed for copyright reasons. dmca.cs.princeton.edu</p> </div>	<div data-bbox="839 304 1408 346">  DMCAInfo @DmcaInfo Follow </div> <p data-bbox="839 367 1408 535">@username Have your tweets ever been taken down for copyright reasons? ©🔥 Answer a few questions for our research, and we'll compensate you \$5 on Paypal—credit you can use for your next cup of coffee</p> <div data-bbox="839 546 1408 924">  <p data-bbox="847 840 1399 924">DMCA Research - Princeton University We'd like your input on a study to help people who have their tweets removed for copyright reasons. dmca.cs.princeton.edu</p> </div>
C. Tag Researcher, No Compensation	D. Tag Researcher, Compensation
<div data-bbox="219 1113 787 1155">  DMCAInfo @DmcaInfo Follow </div> <p data-bbox="219 1176 787 1344">@username Have your tweets ever been taken down for copyright reasons? ©🔥 Answer a few questions for @ohnobackspace's research to help others like you</p> <div data-bbox="219 1354 787 1732">  <p data-bbox="227 1648 779 1732">DMCA Research - Princeton University We'd like your input on a study to help people who have their tweets removed for copyright reasons. dmca.cs.princeton.edu</p> </div>	<div data-bbox="839 1113 1408 1155">  DMCAInfo @DmcaInfo Follow </div> <p data-bbox="839 1176 1408 1354">@username Have your tweets ever been taken down for copyright reasons? ©🔥 Answer a few questions for @ohnobackspace's research, and we'll compensate you \$5 on Paypal—credit you can use for your next cup of coffee</p> <div data-bbox="839 1365 1408 1743">  <p data-bbox="847 1648 1399 1743">DMCA Research - Princeton University We'd like your input on a study to help people who have their tweets removed for copyright reasons. dmca.cs.princeton.edu</p> </div>

Table 6.2. Recruitment message variations

To gather evidence about how much of a difference being more specific about the sampling methods made in the main study period compared to the first recruitment attempt, I logged information about inclusion and exclusion for every account sampled. As shown in Table 6.3, about 71% of accounts sampled were included during the main study period. In total, 1182 recruitment tweets were sent in this time span.

Start– End	Accounts sampled	Accounts excluded language not “en”	Accounts excluded inactive account	Accounts excluded suspended account	Accounts excluded private account	Accounts excluded account not found	Accounts included
3/3/18– 3/16/18	758	5	114	41	15	6	577
3/16/18– 3/20/18	219	2	46	21	4	0	147
3/30/18– 4/5/18	334	3	55	24	11	5	237
4/5/18– 4/11/18	354	4	75	35	9	10	221
Totals:	1665	14	290	121	39	21	1182

Table 6.3. Recruitment attempts during main study period

I also recorded evidence about not only the specificity of the recruitment sampling, but also the different variations in the recruitment messaging that might affect the response rate. Table 6.4 shows the different compensation amounts and recruitment message variations used in each time range, with an estimated page view count. The page view estimation was generated by logging GET requests for the study consent page along with the user agent and query string, and then filtering out user agents that identify users as bots. The query strings identify which tweet

link was used, and I also filtered out the page views that did not follow a recruitment tweet link on Twitter.

Start– End	Accounts included	Compensation	Recruitment message variation	Estimated page views	Average page view per account	Twitter login clicks	Participated
3/3/18– 3/16/18	577	\$3	B	1309	2.27	1	0
3/16/18– 3/20/18	147	\$5	B	462	3.14	0	0
3/30/18– 4/5/18	237	\$0	C	685	2.89	1	0
4/5/18– 4/11/18	221	\$5	D	867	3.92	1	0
Totals:	1182			3323		3	0

Table 6.4. Recruitment and participation

6.2. Recruitment response

As Table 6.4 shows, the study consent page yielded an estimated 3323 total page views during the study period.

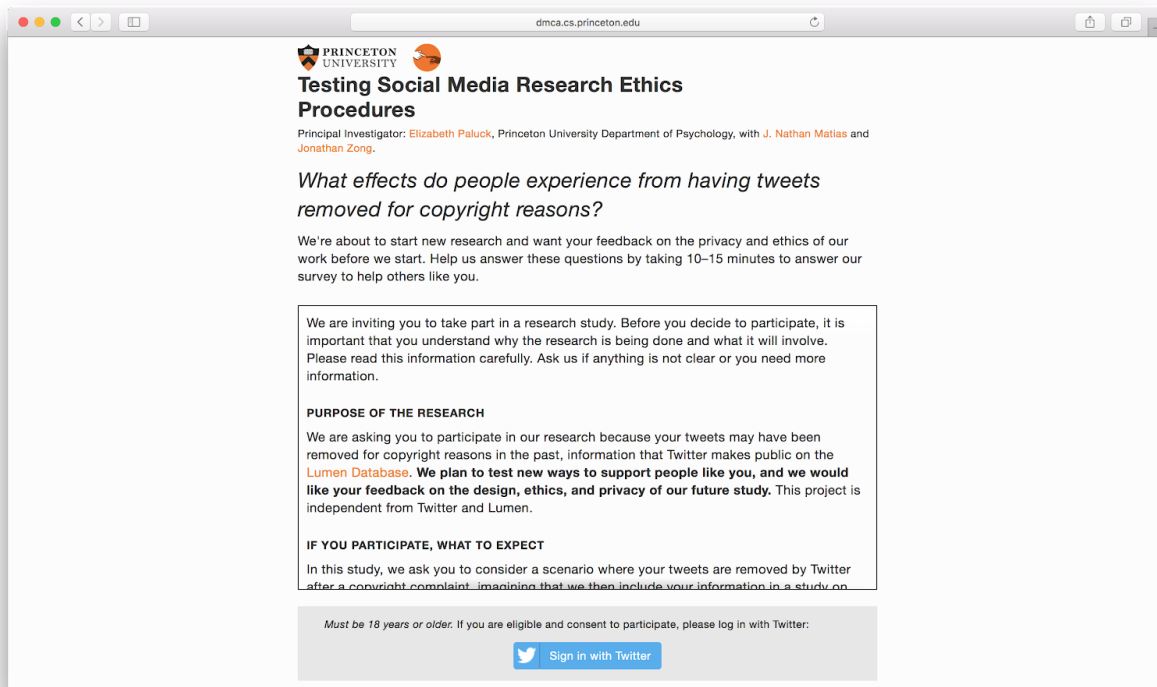


Figure 6.1. Consent page of the survey web application

To consent to take the survey, users must click the Twitter login button to authorize the web application to record their username and some public-facing data which is used to populate fields in the debriefing interface. Table 6.4 shows 3 login button clicks which appear to be from genuine users—they followed shortly after a page view to the consent page which included the recruitment tweet query string. Taken as a percentage of tweets sent, less than a percent of tweets resulted in a click of the Twitter login button on the consent page. Of the 3 login button clicks, none of the users finished authenticating with Twitter to begin the study, meaning they never saw the first question.

6.3. Some hypotheses for low response rate

A prior meta-analysis of different survey modes “showed that on average web surveys yield an 11% lower response rate compared to other modes” [10]. Even then, this study received no responses at all. Why did people not participate? Since this study received no responses, we have no evidence on reasons for non-participation. Here are some possible reasons which might suggest some avenues for future work.

- 1) *Privacy concerns.* Does the fact that clicking through to consent involves authorizing with Twitter deter people from giving over authorization due to privacy concerns? The people in the sample have experienced surveillance through DMCA copyright takedown notices on Twitter. Research on chilling effects due to surveillance shows that knowing about surveillance makes people less likely to read Wikipedia articles about things they think they’re being surveilled about [13]. Perhaps participants are not interested in participating in our copyright related study due to their past experience. It might help to clarify earlier in the process how the Twitter authentication process works and how we use any data that is collected.
- 2) *Personal relevance.* Since the evaluation study asks participants to imagine what the response of like-minded participants of a different study would be, they might not perceive any direct relevance to them personally. The study uses hypothetical placeholder information in the debriefing system. This doesn’t affect the participants’ own data privacy. They are also told that they will not be participants in the future DMCA

debriefing study, so it may not be clear how the research on copyright enforcement effects would directly benefit them.

- 3) *Understanding of the study.* Are participants adequately informed about what the study entails from the consent page? There are a few factors about the design of the consent form that could affect the perception of the study and the participant's state of understanding about it. Adjustments to the content, length, user experience, or appearance of the page might better inform people about the study, its importance, and its value to them.
- 4) *Inconvenient timing.* Perhaps participants received the recruitment message but didn't have time to look at it right away. The survey takes 10–15 minutes to complete, which is information we provide up front on the consent page. Perhaps follow-up reminders to participants would help if participants are receiving the message at an inconvenient moment and therefore not getting around to looking at it.
- 5) *Trust in researchers.* The main recruitment period of this study coincided with public outcry about Cambridge Analytica and its access to Facebook user data. Is participant trust in universities, researchers, and social media platforms' abilities to keep their data safe declining, and might this contribute to the low response rate of this study? To facilitate trust, the recruitment message uses prominent Princeton University branding, but perhaps partnering with an activist group with a more specifically legible focus on copyright issues could help facilitate trust.
- 6) *Unmonitored accounts.* Although the recruitment script filters out inactive accounts, which is what this study calls accounts that have not tweeted in the past week, this

measure is not a perfect proxy for the amount of human attention paid to the account.

Autonomous or semi-autonomous accounts are often referred to as “bots,” although it should be noted that “bot” does not have a single meaning. Even institutional or pseudonymous accounts still have human actors behind them, and all automated scripts are written and maintained by a human actor. These humans still bear legal risk for the content shared by the account and might have reason to respond to our survey. It is possible that due to the nature of our sample, we are not reaching accounts that are as actively monitored as we expect, despite their tweet activity.

- 7) *Financial motivation.* How does the financial compensation for participating in the survey affect the response rate? According to the evidence we collected about page views, the recruitment messages that included a compensation amount received more views per recruitment message sent on average. However, participants may have other motivations that supersede their financial motivation.
- 8) *Credibility of account.* The recruitment messages are clearly automated and delivered by a bot, although they do tag a personal account of a researcher. Does the appearance of the recruitment account have an effect on the participants’ perception of the message? The evidence from the page view logs suggests that users are clicking on the links, so the recruitment message is likely not being dismissed out of hand. However, there also does seem to be a small increase in the page views per recruitment message sent on average for tweets that tag the active, non-pseudonymous account of the researcher in the recruitment message.

These possible reasons people did not participate in the study are not all independent of each other. Some reasons have to do with the effort required to complete the study, while others have to do with the person's willingness to participate in the study. For example, increasing financial motivation might help compensate for the effort it takes to spend time on the study even if the timing is inconvenient, but it may have less effect on participants' willingness to act against their own fear of surveillance or build their trust in researchers. Many of these tightly interconnected factors should be considered when thinking about recruiting users for a study like this one.

7. Future work with debriefing

In this project, I designed a debriefing system for non-consented research. I also designed an evaluation study for that system. Beyond this project, what else might we learn from the debriefing and survey software infrastructure produced in the course of this work?

7.1. Different models of consent

Non-consented research is the use case for the debriefing interface described in this project, but the study used to evaluate the interface still uses informed consent. But since the debriefing interface will be used at a large scale in a non-consented context, how can we explore models of consent that allow us to do the evaluation study and learn about those users while still maintaining the necessary non-consent for them? In his empirical research on research ethics procedures, Desposato outlines a few different models of consent:

- 1) *Individual consent*. In the typical model of consent, each individual must be informed about the experiment procedures and give informed consent for only their own participation in their research.
- 2) *Superset consent*. Researchers employ superset consent by “listing many possible treatments and having the subject consent to the entire set, not knowing which one of them will be used in the experiment,” thereby incorporating a small amount of what could be considered deception at the individual level [3].
- 3) *Representative consent*. In representative consent, participants consent to governance by a representative body made up of people like them. This body is assembled from a subset of the community affected by the research. Participants as a whole are subjected to a decision-making process in which “careful consideration has gone into the decision to offer a particular finding, and that like-minded people, not simply experts, have carefully debated whether that type of information should be offered” [8].

The evaluation study in this project (see Section 5) focuses on this third form of consent, representative consent, which allows us to maintain non-consent for users of the debriefing interface while consulting like-minded representatives in the evaluation of that interface.

7.2. Forecasting

In the evaluation study, we asked a representative sample of people who have received copyright takedown notices to provide feedback on the interface and tell us how they might have used the interface in a hypothetical situation. What if we compared their responses to the actual behavior

of similar users in an actual debriefing situation? A follow-up study of this nature would empirically test the idea of representative consent by asking whether representatives are accurate at forecasting the behavior of others like them.

In this second study, I would recruit English-locale twitter accounts appearing in the Lumen Database into a field experiment that tests the effect on their social media behavior of sending them Twitter messages with information about copyright and artificial intelligence (see Appendix B for a full pre-analysis plan). 4-8 weeks later, I would debrief participants by sending these accounts a link to the debriefing webpage used in the initial forecasting study. They would be assigned to similar variations: 1) whether the debriefing interface includes a graphic of the results, and 2) whether the debriefing interface includes a table of the collected data. The survey would question participants to obtain outcome variables corresponding to the ones in the first study, including information about their past experiences, their decision to opt out of the research, and their views on the risks and benefits of the research. In the analysis, I would compare the outcome variables between the forecasting group and the debriefing group to see if the former can accurately predict the latter.

8. Conclusion

Large-scale online field experiments will only become more widespread as online platforms become increasingly embedded in everyday life. As this happens, fields that engage in behavioral research urgently need to respond to new ethical challenges that arise with this mode of research. The debriefing system proposed by this project aims to establish a norm of post-experiment debriefing for non-consented research, and encode that norm into practice

through reusable software infrastructure. Its design is motivated by the goals of informing users about their participation in research and providing them with control over their data privacy. These goals instantiate the values of informed consent and public accountability that are essential to research ethics.

In running a survey study to evaluate the system, I made a key finding that people did not participate in this research about debriefing despite conventional incentives like compensation. By gathering empirical evidence on participant behavior during recruitment, this project makes progress on understanding the challenges that stand between researchers and the goal of successfully engaging participants in debriefing.

The analysis of non-participation suggests a variety of interdependent factors including privacy concerns, low perceived personal relevance, low understanding of the study from the consent page, inconvenient timing, and others. Conventional incentives like financial compensation address some concerns, like inconvenience, while doing less to change others, like perceptions about privacy.

In discussing all of these findings, the common principle is that as researchers, our thinking about research ethics should always be centered on the participants' perspective. Participant expectations about their activity on online platforms—about how and by whom their data is used, about their inclusion in research when going about their ordinary lives—inevitably frame their encounters with research and data collection online. By treating participants with dignity and care as we seek answers to potentially beneficial research questions, we can preserve the public trust that supports us in our work.

Bibliography

1. Steven Bellman, Eric J. Johnson, and Gerald L. Lohse. 2001. On site: to opt-in or opt-out?: it depends on the question. *Communications of the ACM* 44, 2: 25–27.
2. David J. Cooper. 2014. A Note on Deception in Economic Experiments. *Journal of Wine Economics* 9, 02: 111–114.
3. Scott Desposato. 2014. Ethical Challenges and Some Solutions for Field Experiments. Retrieved April 29, 2018 from <http://www.desposato.org/ethicsfieldexperiments.pdf>.
4. Scott Desposato. 2016. Subjects’ and Scholars’ Views on Experimental Political Science. Retrieved April 29, 2018 from http://swd.ucsd.edu/Scott_Desposato_UCSD/DesposatoEmpiricalEthics.pdf.
5. Casey Fiesler and Nicholas Proferes. 2018. “Participant” Perceptions of Twitter Research Ethics. *Social Media Society* 4, 1: 205630511876336.
6. James Grimmelman. 2015. The Law and Ethics of Experiments on Social Media Users. *Colorado Technology Law Journal* 13.
7. Ralph Hertwig and Andreas Ortmann. 2008. Deception in Experiments: Revisiting the Arguments in Its Defense. *Ethics & behavior* 18, 1: 59–92.
8. Barbara A. Koenig. 2014. Have we asked too much of consent? *The Hastings Center report* 44, 4: 33–34.
9. Robert Kraut, Judith Olson, Mahzarin Banaji, Amy Bruckman, Jeffrey Cohen, and Mick Couper. 2004. Psychological research online: report of Board of Scientific Affairs’ Advisory Group on the Conduct of Research on the Internet. *The American psychologist* 59, 2: 105–117.
10. Katja Lozar Manfreda, Jernej Berzelak, Vasja Vehovar, Michael Bosnjak, and Iris Haas. 2008. Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research* 50, 1: 79–104.
11. J. Nathan Matias. How Data Science and Open Science are Transforming Research Ethics: Edward Freeland at CITP. Retrieved May 2, 2018 from <https://freedom-to-tinker.com/2018/02/07/how-data-science-and-open-science-are-transforming-research-ethics-edward-freeland-at-citp/>.
12. J. Nathan Matias. 2018.
13. Jon Penney. 2016. Chilling Effects: Online Surveillance and Wikipedia Use. *Berkeley Technology Law Journal* 31, 1: 117.
14. 2016. 45 CFR 46. Retrieved May 7, 2018 from <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html>.

Appendix

A. Code for the debriefing system

The complete code for the debriefing system can be found on GitHub at:

<https://github.com/jonathanzong/dmca>

The latest commit hash at time of publication is:

6a3961a9dba2ab288932713f64914b0df2ab4fd2

The state of the repository at this commit can be viewed here:

<https://github.com/jonathanzong/dmca/tree/6a3961a9dba2ab288932713f64914b0df2ab4fd2>

B. Forecasting and debriefing study pre-analysis plan

The draft pre-analysis plan for the full forecasting and debriefing study referenced in Section 7 is attached in its entirety beginning on the next page.

Estimating Effects of Research Debriefing Interface Designs on Research Participant Perceptions and Behavior Toward Online Research

Jonathan Zong

Introduction

As behavioral experimentation becomes more widespread in society through online platforms, we need new ways to manage the ethics and accountability of that research. Since this research is delivered digitally, we can develop novel technologies for managing large-scale research ethics. Because models of consent and accountability in research ethics involve communicating complex ideas to the public, advances in user interfaces for managing participation in research can contribute to novel approaches in research ethics.

For example, in large-scale experiments online, due to practical concerns obtaining informed consent from the entire population is not always possible. Under the Common Rule, IRB can waive requirement for signed consent form by the following criteria: the study must have minimal risk, obtaining informed consent must be impractical, and there must be a post-experiment debriefing.

Debriefing is a procedure in experiments involving human subjects wherein, after the experiment has concluded, participants are provided with information about the experiment and the data that was collected in the process. The procedure serves an important ethical purpose by giving the participants an opportunity to clarify their involvement, ask questions, or opt-out; this is especially important in experiments where there was any form of deception or informed consent was not obtained beforehand. Because successful debriefing requires people to understand the experiment, novel user interface approaches may improve the debriefing process.

Do variations in what kinds of user interface elements—like tables, charts—we use to present information in a debriefing interface have any effect on the likelihood of participants in research to opt-out of data collection or adopt certain perceptions about the value of the research or how it is conducted? It's possible that more transparency into the research process might help participants calibrate their understanding of the risks and benefits of being included in research. It's also possible that they might be discouraged from participating due to concerns about privacy and data collection. This experiment tests the effect of variations in the debriefing interface on participants' opt-out behavior and attitudes about the research process.

Study Procedure

This study has two parts. In the first part of this study, we ask Twitter users who have received DMCA copyright notices in the past to give feedback on a web interface for debriefing participants in field experiments. We will also survey them about research ethics and their choice to opt out of the research. In the second part of this study, we debrief a new set of participants from the same group and compare the forecasts of the first group to the responses and actions of the second.

- Forecasting study:
 - **Recruit** Twitter accounts that have received copyright notices in the two months prior to the beginning of the pilot. These notices are a matter of public record in the Lumen database. We sample from this population because they share the experience of receiving a copyright notice, together with the study population we want to forecast for the second study.
 - Participants will be included in this study if:
 - if they appear in the Lumen database of Twitter DMCA takedown notice
 - if the database records that they received a DMCA takedown notice within the past two months
 - if we identify links in the notice to this participant's Twitter account
 - if we can successfully identify that Twitter account via the Twitter API
 - if Twitter reports that the account has a "en" language (which is a proxy for locale)
 - Participants are recruited by @-messages sent to their twitter account, with a link to the debriefing interface and survey
 - Participants are compensated for completing the survey
 - The intervention:
 - Asks participants to imagine that they had been part of a field experiment
 - Shows participants a debriefing interface
 - Group participants into a stratified sample of people whose content was removed for copyright reasons, and those whose content was permitted to remain on Twitter, to ensure balance across experiment arms between both groups.
 - Randomly assigns participants to variations
 - Whether they are assigned to the **control group** of the imagined experiment or not
 - Whether the debriefing interface includes a **graphic** of the results
 - Whether the debriefing interface includes a **table** of the collected data

- Throughout the debriefing experience, we will survey participants to obtain the outcome variables. These variables include information about their past experiences, their forecasted behaviors, and their views on the risks and benefits of the research.
- Upon completing the survey, participants will be compensated for their participation
- Debriefing study:
 - **Recruit** English-locale twitter accounts appearing in the Lumen Database into a field experiment that tests the effect on their social media behavior of sending them Twitter messages with information about copyright and artificial intelligence (see pre-analysis plan).
 - 4-8 weeks later, **debrief** participants by sending these accounts a link to the debriefing webpage used in the Forecasting Study. These participants will not be compensated.
 - Randomly assign participants to the following variations
 - Whether the debriefing interface includes a **graphic** of the results
 - Whether the debriefing interface includes a **table** of the collected data
 - Survey participants to obtain the outcome variables. These variables include information about their past experiences, their decision to opt out of the research, and their views on the risks and benefits of the research
- Comparison between Forecasting and Debriefing
 - In this analysis, we will compare the outcome variables between the forecasting group and the debriefing group, as specified below

Outcome Variables

The following variables will be used to estimate the effect of debriefing interface variations on participants' behaviors and attitudes about data privacy and inclusion in research studies. The dataframe contains one row per participant. Its columns are the outcome and other variables described below. Some outcome variables are specific to the forecasting study or the debriefing study, while some outcomes encode survey questions which are shared between the two studies. The analysis will include a comparison of effects between the two studies.

Forecasting Study Outcomes

Click Debrief Tweet

In this five-point likert survey question on a scale from -5 to 5 (see Supplementary Materials), we ask about the hypothetical debriefing tweet that the participant would receive from us notifying them that they had been in an experiment: "How likely would you be to click the link?"

We use the answer from the forecasting group to estimate the real click-through rate of the debriefing group.

```
forecasting.participant$click.tweet
```

Would Opt Out

In this five-point likert survey question on a scale from -2 to 2 (see Supplementary Materials), we ask about the likelihood that the participant would delete their data from the hypothetical study using the debriefing interface we show them: "In the debriefing webpage, we gave you a chance to have your data deleted from the study. How likely would you be to click the button to delete your data?" We use the answer from the forecasting group to estimate the real opt-out rate of the debriefing group.

```
forecasting.participant$would.optout
```

Vote on Study

In this three-point ordinal survey question on a scale from -1 to 1 (see Supplementary Materials), we ask about how participants would vote if they could vote on whether the hypothetical study proceeds: "If you could vote on whether this study should happen, how would you vote?" We use the answer from the forecasting group to estimate the response of the debriefing group.

```
forecasting.participant$vote.study
```

Debriefing Study Outcomes

Click Debrief Tweet

This binary variable represents whether or not the user does not click (0) or does click (1) on the link in the debrief tweet to view the debriefing interface.

```
debriefing.recruits$click.tweet
```

Opts Out of Debriefing Study

This binary variable represents whether or not the user chooses to remain in the research (0) or opt-out of data collection (1) using the debriefing interface.

```
debriefing.participant$opted.out
```

Outcomes Common to Both Forecasting and Debriefing

Society Benefit

In this five-point likert survey question on a scale from -2 to 2 (see Supplementary Materials), we ask about the participant's assessment of the magnitude and direction of potential benefits to society in the copyright study: "How beneficial to society would it be to learn whether copyright enforcement affects speech on Twitter?" We use the answer from the forecasting group to estimate the response of the debriefing group.

```
forecasting.participant$society.benefit  
debriefing.participant$society.benefit
```

Personal Benefit

In this five-point likert survey question on a scale from -2 to 2 (see Supplementary Materials), we ask about the participant's assessment of the magnitude and direction of potential benefits to themselves personally in the copyright study: "How much might research on the effects of online copyright enforcement benefit you personally?" We use the answer from the forecasting group to estimate the response of the debriefing group.

```
forecasting.participant$personal.benefit  
debriefing.participant$personal.benefit
```

Surprised by Data Collection

In this four-point ordinal survey question on a scale from 0 to 3 (see Supplementary Materials), we ask about the participant's surprise that their public Twitter behavior could be observed in the manner described in the copyright study: "Suppose you learned that you were one of the participants in this study. How surprised are you that we are able to collect this information about your public Twitter behavior?" We use the answer from the forecasting group to estimate the response of the debriefing group.

```
forecasting.participant$collection.surprised  
debriefing.participant$collection.surprised
```

Glad Included in Study

In this three-point ordinal survey question on a scale from -1 to 1 (see Supplementary Materials), we ask about whether the participant would feel positive, negative, or neutral about their involvement in the copyright study: "Which of the following best describes how you would feel about being included in the study?" We use the answer from the forecasting group to estimate the response of the debriefing group.

```
forecasting.participant$glad.included  
debriefing.participant$glad.included
```

Share Results

In this three-point ordinal survey question on a scale from 0 to 2 (see Supplementary Materials), we ask to what extent the participant would be interested in sharing the results of copyright study: "If we sent you what we learn, what best describes how you might share the results of this research online with others?" We use the answer from the forecasting group to estimate the response of the debriefing group.

```
forecasting.participant$share.results  
debriefing.participant$share.results
```

Improve Debrief

In this freeform text survey question, we ask about what changes participants might suggest for the debriefing interface: "If we could make the research debriefing webpage different, what would you change? (optional)."

```
forecasting.participant$improve.debrief  
debriefing.participant$improve.debrief
```

Other Variables Important to Experiment Procedures and Analysis

The following variables are non-outcome variables (not dependent on the condition variables described in the next section). They are used to record information assigning participants into groups relevant to the analysis.

Content Removed

In this binary survey question, we ask about the DMCA copyright takedown notice that the participant received: "When this happened, did Twitter remove your Tweet or media?" We use the answer to assign the participant to a group based on their answer to this question, and randomly assign conditions within each group.

```
forecasting.participant$content.removed  
debriefing.participant$content.removed
```

Sampling and Conditions

Population / sampling method

The pilot study population includes Twitter users who have received Lumen notices in the past 60 days from the start of the study, who have their language set to 'en', and have tweeted at least once in the week before the recruitment attempt. Recruitment used a randomized sample from this population. The sampling method was stratified sampling, with two possible strata: "Removed" and "Not Removed", referring to whether or not the Tweet identified in the copyright notice was removed by Twitter.

Conditions

The pilot study has 3 binary condition variables, for a total of $2^3 = 8$ conditions.

- in_control_group
 - was the participant assigned to the control group in the hypothetical study?
- show_table
 - was the participant shown their collected data in a table?
- show_visualization
 - was the participant shown a visualization of the results?

Code for Estimation of Treatment Effects

In the analysis, we use a dataframe where each row is a participant and columns contain the condition variables and outcome variables relevant to the analysis. For survey questions which are likert or ordinal variables, we use a linear regression model to estimate the average treatment effect for participants. For opt-out, which is a binary outcome, we use a logistic regression model. The decision rule will be $\alpha=0.05$. Results will be adjusted for multiple comparisons done within the dataset being analyzed. For example, we conduct 12 statistical tests on the forecasting study and will adjust the results using the Bonferroni method for 12 comparisons.

Effects Within The Forecasting Study

Effect on Forecasted Likelihood to Opt Out

We expect the following outcomes:

Hypothetically being in the control group will decrease a person's reported likelihood to opt out, compared to being in the treatment group.

```
lm(would.optout ~ in_control_group,  
   data=forecasting.participants)
```

Seeing a table with the data collected about them will decrease a person's reported likelihood to opt out, compared to not seeing the table

```
lm(would.optout ~ show_table, data=forecasting.participants)
```

Seeing a graphic illustrating the person's own observed behavior will decrease a person's reported likelihood to opt out, compared to not seeing the graphic

```
lm(would.optout ~ show_visualization,  
   data=forecasting.participants)
```

Effect on Forecasted Perceived Benefit to Society

We expect the following outcomes:

Hypothetically being in the control group will decrease a person's reported assessment of the study's benefit to society, compared to being in the treatment group.

```
lm(society.benefit ~ in_control_group,  
   data=forecasting.participants)
```

Seeing a table with the data collected about them will increase a person's reported assessment of the study's benefit to society, compared to not seeing the table

```
lm(society.benefit ~ show_table, data=forecasting.participants)
```

Seeing a graphic illustrating the person's own observed behavior will increase a person's reported assessment of the study's benefit to society, compared to not seeing the graphic

```
lm(society.benefit ~ show_visualization,  
   data=forecasting.participants)
```

Effect on Forecasted Perceived Personal Benefit

We expect the following outcomes:

Hypothetically being in the control group will decrease a person's reported assessment of the study's personal benefit to them, compared to being in the treatment group.

```
lm(personal.benefit ~ in_control_group,  
data=forecasting.participants)
```

Seeing a table with the data collected about them will increase a person's reported assessment of the study's personal benefit to them, compared to not seeing the table

```
lm(personal.benefit ~ show_table,  
data=forecasting.participants)
```

Seeing a graphic illustrating the person's own observed behavior will increase a person's reported assessment of the study's personal benefit to them, compared to not seeing the graphic

```
lm(personal.benefit ~ show_visualization,  
data=forecasting.participants)
```

Effect on Forecasted Surprise at Data Collection

We expect the following outcomes:

Hypothetically being in the control group will increase a person's surprise that the data collection was possible, compared to being in the treatment group.

```
lm(collection.surprised ~ in_control_group,  
data=forecasting.participants)
```

Seeing a table with the data collected about them will increase a person's surprise that the data collection was possible, compared to not seeing the table

```
lm(collection.surprised ~ show_table,  
data=forecasting.participants)
```

Seeing a graphic illustrating the person's own observed behavior will increase a person's surprise that the data collection was possible, compared to not seeing the graphic

```
lm(collection.surprised ~ show_visualization,  
data=forecasting.participants)
```

Effects Within The Debriefing Study

Effect on the Decision to Opt Out

We expect the following outcomes:

Hypothetically being in the control group will decrease a person's probability to opt out of the study, compared to being in the treatment group.

```
glm(optimized.out ~ in_control_group, family=binomial,  
    data=debriefing.participants)
```

Seeing a table with the data collected about them will decrease a person's probability to opt out of the study, compared to being in the treatment group.

```
glm(optimized.out ~ show_table, family=binomial,  
    data=debriefing.participants)
```

Seeing a graphic illustrating the person's own observed behavior will decrease a person's probability to opt out of the study, compared to being in the treatment group.

```
glm(optimized.out ~ show_visualization, family=binomial,  
    data=debriefing.participants)
```

Effect on Perceived Benefit to Society

We expect the following outcomes:

Hypothetically being in the control group will decrease a person's reported assessment of the study's benefit to society, compared to being in the treatment group.

```
lm(society.benefit ~ in_control_group,  
    data=debriefing.participants)
```

Seeing a table with the data collected about them will increase a person's reported assessment of the study's benefit to society, compared to not seeing the table

```
lm(society.benefit ~ show_table, data=debriefing.participants)
```

Seeing a graphic illustrating the person's own observed behavior will increase a person's reported assessment of the study's benefit to society, compared to not seeing the graphic


```
lm(society.benefit ~ show_visualization,  
data=debriefing.participants)
```

Effect on Perceived Personal Benefit

We expect the following outcomes:

Hypothetically being in the control group will decrease a person's reported assessment of the study's personal benefit to them, compared to being in the treatment group.

```
lm(personal.benefit ~ in_control_group,  
data=debriefing.participants)
```

Seeing a table with the data collected about them will increase a person's reported assessment of the study's personal benefit to them, compared to not seeing the table

```
lm(personal.benefit ~ show_table, data=debriefing.participants)
```

Seeing a graphic illustrating the person's own observed behavior will increase a person's reported assessment of the study's personal benefit to them, compared to not seeing the graphic

```
lm(personal.benefit ~ show_visualization,  
data=debriefing.participants)
```

Effect on Surprise at Data Collection

We expect the following outcomes:

Hypothetically being in the control group will increase a person's surprise that the data collection was possible, compared to being in the treatment group.

```
lm(collection.surprised ~ in_control_group,  
data=debriefing.participants)
```

Seeing a table with the data collected about them will increase a person's surprise that the data collection was possible, compared to not seeing the table

```
lm(collection.surprised ~ show_table,  
data=debriefing.participants)
```

Seeing a graphic illustrating the person's own observed behavior will increase a person's surprise that the data collection was possible, compared to not seeing the graphic

```
lm(collection.surprised ~ show_visualization,  
data=debriefing.participants)
```

Comparing Effects Between the Two Studies

This section is left for future work.