Ethical-Response Survey Report: Fall 2014

WORKING PAPER — EXPECT CHANGES

Stuart Schechter
Microsoft Research

Cristian Bravo-Lillo
Universidad de Chile

November 2, 2014

Abstract

We update the ethical-response survey we published in July [11] to broaden its reach in two dimensions. In addition to surveying workers on Amazon's Mechanical Turk, we also reached out to juror candidates who had been summoned to serve at the King County Superior Court in Seattle, WA. In addition to five experimental scenarios we examined in prior surveys, we added seven new scenarios: two designed to serve as baselines of innocuousness and concern; two censorship-detection experiments that the Internet Measurement Conference refused to publish on ethical grounds, two human-subjects experiments for which researchers requested feedback, and one experiment (by OKCupid) that received attention in the press.

1 Introduction

This is the first of what we hope will be many periodic updates to our ethical-response survey, in which we present respondents with experimental scenarios and ask two questions designed to gauge whether the experiments are ethical:

1. If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?

2. Do you believe the researchers should be allowed to proceed with this experiment?

While the long-term goal is to use the survey to help researchers anticipate the ethical implications of the planned experiments *before* performing them, we have focused initially on scenarios based on previously-performed experiments.

In this update, we examine seven new experimental scenarios in addition to the five covered in our previous survey. We also reach out to a second respondent pool, juror candidates in King County WA, to augment the workers on Mechanical Turk who constituted the participant pool for our previous survey.

2 New scenarios

After publishing our first such survey [11] to explore five experimental scenarios, we have invited researchers to submit scenarios for experiments that had previously conducted or planned to conduct. We also added two new scenarios to act as baselines for innocuous research and research that received a level of disapproval one would want to avoid.

2.1 Baseline scenarios

While our prior surveys were instructive in allowing us to make relative comparisons between experimental scenarios, we failed to provide thresholds at which one could feel confident allowing or rejecting a study. We thus added two baseline studies to provide a threshold at which one could conclude a study was *innocuous* and another at which one could conclude it was so *objectionable* as to be confident disallowing it.

Favorite color

In our *innocuous* baseline scenario, we describe an experiment in which researchers ask workers on Amazon's Mechanical Turk for their gender and favorite color, to determine whether favorite color varies by gender. The researchers publish anonymous aggregate statistics but do not collect any personal information.

Urinal observation

In our *objectionable* baseline scenario, we describe an experiment published in 1976 by Middlemist *et al.* [8], in which the researchers surreptitiously observed the lower torso and urine stream of men at urinals in a university restroom. The purpose of the study was to measure whether the delay at the start of urination and the duration of urination changed due to proximity to others. The response to that study is one that researchers, and their institutions, would want to avoid.

2.2 Censorship-detection experiments

In July, the 2014 Internet Measurement Conference rejected two independent papers on censorship-detection systems, citing ethical objections as the primary factor in their decision. Both papers described systems for detecting censorship indirectly, using computers in foreign countries that were neither owned nor operated by the researchers.

Censorship-detection webpage

The first rejected paper described a system for creating web pages that, when loaded, would request images and other content from servers that the researchers suspected might be censored by the country in which the user's computer was located. For the pilot study described in the paper, the authors limited the set of potentially-censored websites they tested to large The web page would attempt to load this content without notifying or receiving consent from the user, then report back to the server to indicate whether the browser was able to load the content.

The researchers reasoned that many web pages already load analytics and advertising content without notifying or receiving users' consent, and so their code was acting within the norms of website behavior and not doing anything users didn't (or shouldn't) already expect. In the words of their project webpage[2] (not quoted in our scenario):

> Countless sites already load resources from numerous third parties and record detailed measurements about users' browsing behavior without informing them or obtaining their consent. (For example, this paper studies the pervasiveness of third-party Web tracking.) Rather than loading third party resources for commercial purposes (e.g., to track users or build consumer advertising profiles), Encore uses the same mechanisms "for good" to measure Internet censorship.

On the other hand, few advertising firms attempt to download content that they expect to be forbidden in the user's jurisdiction. The researchers tested their system by deploying it on their personal home pages and collecting data from the browsers of users who loaded their webpages.

A number of additional facts regarding to the practice of ethical review and regulation, which are salient to the paper, were not salient to the scenario and thus did not appear in our survey. First, before pilot testing, the authors contacted their IRB chair by phone and report being were told that the experiment did not require review from the IRB. For a future submission, the lead author contacted the IRB chair again, by email and then in person, and received written confirmation that the IRB did not want to review it [3] (copied verbatim in Appendix A).

Second, the authors report restricting the set of potentially-censored sites to reduce the potential risk to participants — focusing on websites already likely to have content embedded on many websites popular in those jurisdictions and that governments would presumably find least objectionable [6]. (Our scenario simply referred to "potentially-censored sites" without discussing how objectionable the government might find each site.)

After receiving their rejection from the Internet Measurement Conference on ethical grounds, the authors sought out ethical guidance from the research community, courageously[1] offered up their experiment as a case study in research ethics, and shared the experiment with us so we could include it in our study. The senior author worked with us to ensure that the summary of the experiment used in our survey was accurate. The authors later made the research available to the public via arXive [1].

Censorship-detection scan

The second rejected paper presents system for testing China's national firewall for holes that would allow citizens to access the TOR anonymity service—a service that makes it harder for governments to monitor citizens' web use and which provides access to parts of the web that are blocked by China. To look for holes, the researchers identified a random set of computers located in China (computers they did not own nor have consent to use) and sent messages (packets) that tricked these computers into replying with

---

[1]Conflict of interest disclosure: our lead author Stuart Schechter, has known Professor Feamster, the senior author of the paper, since 2002, when Dr. Feamster attended graduate school with, and shared an office with, Dr. Schechter's spouse.

messages to the Tor service. They tested only connectivity, and did not cause any forbidden content to be sent or received. The researchers also used the same trick to test whether computers from the Tor service can send messages to the Chinese computers.

Whereas the other censorship experiment required some human interaction to make the system work, and so there is ambiguity about whether an IRB submission was required under the common rule [12], this experiment appears unambiguously exempt from human-subjects review. The experiment measured only the behavior of the network and the computers on the network and there was no interaction with humans or measurements of human behavior. If all of human kind ceased to exist the moment the experiment began it would still be expected to return the same result.

Yet, despite the fact that this paper clearly falls outside the jurisdiction of the common rule, which applies only to human-subjects experiments, the experiment posed a risk to humans: China might detect the messages that computers were tricked to sending to the forbidden Tor service, conclude that the owners of those computers were trying to use Tor and thereby breaking the law, and punish those owners for crimes they had never committed. However, the researchers knew of no evidence that the government of China had ever detected citizens using Tor or punished them for doing so.

After receiving their rejection from the Internet Measurement Conference on ethical grounds, the researchers received a request from us, routed through the IMC program committee to protect their anonymity. They generously and courageously shared a copy of their submission and helped us to craft a description of the experiment so that we could include it in our survey. They too have now made their research available to the public via arXive [5].

2.3 Other recent experiments

Dating matches

This scenario was based on an experiment that Christian Rudder [10], in a blog post in late July 2014, revealed he had performed on users of OKCupid (where he is CEO). In our scenario, researchers at "an online dating site" want to determine whether their matching algorithm is effective in identifying individuals who will want to communicate, or whether the correlation is due to the power of suggestion. To do so they show some users false good matches (bad matches presented as good matches) and other users false bad matches to see whether suggestion alone is

responsible for increased communication. As revelations about the experiment received significant press coverage, we not only asked our standard question about whether respondents had heard about experiments like this one in the past, but asked specifically about the experiment and naming OKCupid at the end of the survey. We are grateful to Christian Rudder who, on a visit to Microsoft to discuss his new book on September 18 2014, generously took the time to proofread our scenario and correct a number of inaccuracies before we went to final draft.

Security-feature adoption

This scenario describes an experiment [4] in which researchers at Facebook ("a social network" in our scenario) tested whether users might be more likely to adopt a security protection if they knew that many of their friends had. The researchers selected 50,000 users at random and assigned them to treatment groups. One treatment group received an announcement of the security feature at the top of their news feed with no information about how many of their friends were already using it. One treatment group was shown how many of their friends are already using the protection. Another group was shown the percent of their friends already using it. The researchers excluded from the study those users who have fewer than ten friends using protection, to prevent users from inferring which friends were and were not using it.

We first added this scenario at the request of its lead author, Sauvik Das[2] and the consent of all co-authors. The experiment was accepted for publication at the 2014 ACM Conference on Computer and Communications Security (CCS).

Delayed search

In this experiment, experimenters at Microsoft's[3] Bing search engine ("a search engine" in our scenario) wanted to measure how much of an impact latency had on customer satisfaction, retention, and other metrics that they collect on a day-to-day basis [7]. They added a small delay of one tenth of a second to searches made by 10% of users and one quarter of a second to another 10% of users (in the scenario we simplified this to "one to two tenths of a second").

---

[2]Conflict-of-interest disclosure: Sauvik Das, the lead author of the paper, was an intern at Microsoft Research working for our lead author, Stuart Schechter, at the time. Das's co-authors included Adam Kramer, the lead author of Facebook's emotional contagion study, which we also examine in our survey.

[3]Conflict-of-interest disclosure: our lead author is employed by Microsoft.

Users were not notified that they were part of the experiment and did not have a chance to opt out. The experimenters reasoned that the slowdown was sufficiently small to cause minimal harm and would not reduce quality-of-service below acceptable levels for a free offering.

We added this scenario with the cooperation of Ronny Kohavi from Microsoft's Bing team.

3 Experimental procedure

We offered our survey in return for a $5.00 to two populations.

On Friday September 19, one researcher (the lead author) stood outside the King County Courthouse handing envelopes to US Citizens residing in King County who were arriving for jury service. Each postage-paid envelope was addressed to the researcher's office at Microsoft, had a $5.00 bill taped to the outside, and contained a copy of the survey with the scenario ordering randomized. The researcher coordinated this activity with the courthouse and informed prospective participants that he was not affiliated with the court, that they were not obligated to accept the survey, and that they could keep the money regardless of whether they completed the survey. If unsure whether individuals arriving at the courthouse were among those summoned for jury duty, the researcher asked to see their summons. The researcher brought 105 envelopes and handed out all of them between 7:00AM and 8:00AM, the time at which jurors were required to be at the courthouse. The county had summoned over 400 juror candidates that day and, once inside, candidates were required to turn off all electronic devices and wait for over an hour for instructions—providing candidates with little else to occupy their time.

Three days later, on Monday September 22, we offered the survey as a $5.00 Human Intelligence Task (HIT) on Amazon's Mechanical Turk crowdsourcing service, requiring that workers come from the US. Our online survey and offline survey were generated from the same source file to ensure consistency between the two. We did require workers on Mechanical Turk to complete the survey in order to receive compensation, though if they preferred not to answer most questions they could choose not to answer. The surveys differed only in their instructions regarding the mechanics of the survey (e.g., respondents were asked to circle answers on the paper survey) and the survey environment (e.g., respondents in the jury sample were instructed on how to mail their survey). Using an online survey, we were able to enforce that participants respond to each of the multiple choice questions, whereas some participants in the jury sample missed individual questions or skipped pages. (The resulting inconsistencies in the number of participants who answered different questions can be seen in the data in our results section.)

We have attached to the appendix an instance of the jury survey for participant 1. Instances printed for other participants would have different random ordering for the scenarios.

3.1 Questions for each scenario

As with prior implementations of our ethical-response survey, our analysis focuses on two key questions that followed each scenarios.

Q1 measured respondents' *concern* for those participating in the experiment. We asked: "If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?"

We asked respondents about someone they care about, as opposed to themselves, because they might be more comfortable imagining others to be vulnerable and needing protection, whereas they might not want to admit being vulnerable themselves. We provided the option to respond "Yes", "I have no preference", or "No". We say that participants expressed *concern* for participants if they answered "No".

We always asked about concern for participants first to give respondents a chance to humanize potential participants, and to think about the consequences of the experiment on them, before we asked Q2: "Do you believe the researchers should be allowed to proceed with this experiment?"

We offered three response options for Q2, again ordered from most approving to least approving with the first option being "Yes" (on the left) and the last "No" (on the right). In between these options we included a hedge, "Yes, but with caution", for respondents who did not want to disapprove of a experiment but feared that an unambiguous "yes" would absolve researchers of all other ethical responsibilities. (In prior surveys we had a fourth option, "I'm not sure", which we have removed so that respondents would have to commit to an answer.) We say that respondents exhibited *disapproval* of an experiment if they responded "No".

For each of these first two questions, we gave respondents a free-response field in which they could optionally explain their answers.

We also asked respondents "Are you aware of having ever participated in such a study?" and "Are you aware of a study like this one having been performed

4

Question	Workers on Mechanical Turk	Juror Candidates
In what year were you born? *(median values shown)*	1984 (n=98)	1965 (n=18)
Please indicate whether your annual household income is above $53,000.	34/99 (34%)	24/31 (77%)
Have you ever participated in a study that involved deception?	36/99 (36%)	2/31 (6%)
Have you ever purchased goods advertised via an unsolicited marketing email?	5/99 (5%)	2/30 (7%)

Table 1: Our two sample populations had median ages two decades apart and incomes that fell on different sides of the national median. A much greater fraction of workers on Mechanical Turk had previously participated in deception experiments.

by researchers in the past? (For example, have you heard about it in the news or learned about it in a class?)" The response options, from left to right, were "Yes" and "No".

3.2 Closing questions

After respondents completed the twelve experimental scenarios, we asked for their year of birth, gender, occupation (free response), whether they had ever purchased goods advertised via unsolicited email (for insight into an experiment about spam), whether they had participated in a study involving deception, and whether they had heard about "Facebook's 'mood' study" or OKCupid's study.

4 Results

Of the 105 workers on Mechanical Turk who completed our survey, 99 respondents spent more than 12 minutes doing so (one minute per scenario) yielding a 94% true completion rate. We paid all 105, but discarded the responses of the six who spent less than 12 minutes.

Of the 105 juror candidates who accepted a survey envelope with the $5 bill attached, we received 32 responses, yielding a 30% completion rate.

Previous studies using jury pools as experimental participants have had higher completion rates when researchers were allowed to operate within the courthouse with the full cooperation of the court [9]. The King County WA courthouse cooperated in so far as to explain where we could legally position ourselves outside the courthouse, so as to keep us from violating court rules, and by identifying the location of a mailbox within the courthouse from which participants could be instructed to mail their surveys. However, the court did not allow us into the courthouse out of fear of the precedent that would be set by allowing access for "corporate" research. Thus, while reaching out to candidate jurors can potentially provide a participant sample that is more diverse than students or workers on Mechanical Turk, we must recognize two potential sources of sample bias beyond the expected geographic bias: only a minority of the juror candidates who came to the courthouse accepted our survey and only 30% of those who accepted mailed in a completed survey.

We present respondents' answers to our two primary ethical-response questions in Table 2, ordered by the fraction of workers on Mechanical Turk who responded that each experimental scenario should not be allowed to proceed (a "no" on Q2). We do not see much difference in the ordering of scenarios between the workers on Mechanical Turk and the (much smaller) sample of juror candidates.

One comment from a juror-candidate participant (J001) indicated an ethical concern with the innocuous baseline study: "This violates the privacy of an individual's opinion – also could lead to inaccurate portrayals in the media" (the same argument could be used to object to any survey). Most other comments from the small number of participants who objected to the innocuous baseline, measuring favorite color, did not indicate ethical concerns so much as a sense that money would be wasted studying the question. This can be seen as a fault of our scenario choice, as it caused false reporting of ethical concern. On the other hand, the baseline helps to show that one should expect a certain number of participants to respond that even innocuous are unethical, and expect others to argue that experiments they find harmless shouldn't proceed for other reasons, and so to account for the existence of these opinions in the population when setting a threshold of benign response that one might aspire to reach.

The OKCupid study has scores similar to roughly on par with Facebook's emotional contagion study.

The two scenarios based on experiments that researchers had requested that we examine, the one using social cues to encourage adoption of security features and the one that slowed down search engine

Scenario	Workers on Mechanical Turk			Juror Candidates		
	no	*indifferent*	*yes*	*no*	*indifferent*	*yes*
Favorite color	5/99 (**5%**)	15/99 15%	79/99 80%	2/31 (**6%**)	18/31 58%	11/31 35%
Spoofed-warning deception	12/99 (**12%**)	31/99 31%	56/99 57%	9/32 (**28%**)	13/32 41%	10/32 31%
Delayed search	14/99 (**14%**)	40/99 40%	45/99 45%	8/32 (**25%**)	20/32 62%	4/32 12%
Password-dialog spoofing	22/99 (**22%**)	22/99 22%	55/99 56%	11/32 (**34%**)	8/32 25%	13/32 41%
Security-feature adoption	20/99 (**20%**)	38/99 38%	41/99 41%	8/32 (**25%**)	16/32 50%	8/32 25%
Spam takeover	40/99 (**40%**)	32/99 32%	27/99 27%	15/32 (**47%**)	12/32 38%	5/32 16%
Social phishing	39/99 (**39%**)	26/99 26%	34/99 34%	13/32 (**41%**)	11/32 34%	7/32 22%
Dating matches – *unaware*	41/88 (**47%**)	23/88 26%	24/88 27%	10/26 (**38%**)	13/26 50%	3/26 12%
Dating matches – *all*	47/99 (**47%**)	25/99 25%	27/99 27%	12/31 (**39%**)	15/31 48%	4/31 13%
Emotional contagion – *unaware*	34/76 (**45%**)	16/76 21%	26/76 34%	12/24 (**50%**)	8/24 33%	4/24 17%
Emotional contagion – *all*	51/99 (**52%**)	20/99 20%	28/99 28%	18/32 (**56%**)	9/32 28%	5/32 16%
Censorship-detection webpage	71/99 (**72%**)	18/99 18%	10/99 10%	21/32 (**66%**)	9/32 28%	2/32 6%
Censorship-detection scan	72/99 (**73%**)	15/99 15%	12/99 12%	22/32 (**69%**)	5/32 16%	5/32 16%
Urinal observation	68/99 (**69%**)	19/99 19%	12/99 12%	23/32 (**72%**)	7/32 22%	2/32 6%

(a) **Concern for participants.** Q1: "If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?"

Scenario	Workers on Mechanical Turk			Juror Candidates		
	no	*indifferent*	*yes*	*no*	*caution*	*yes*
Favorite color	3/99 (**3%**)	3/99 3%	93/99 94%	4/31 (**13%**)	6/31 19%	21/31 68%
Spoofed-warning deception	8/99 (**8%**)	19/99 19%	72/99 73%	7/32 (**22%**)	9/32 28%	16/32 50%
Delayed search	11/99 (**11%**)	22/99 22%	66/99 67%	7/32 (**22%**)	10/32 31%	15/32 47%
Password-dialog spoofing	14/99 (**14%**)	34/99 34%	51/99 52%	7/32 (**22%**)	15/32 47%	10/32 31%
Security-feature adoption	16/99 (**16%**)	22/99 22%	61/99 62%	9/32 (**28%**)	8/32 25%	15/32 47%
Spam takeover	23/99 (**23%**)	33/99 33%	43/99 43%	10/32 (**31%**)	16/32 50%	6/32 19%
Social phishing	28/99 (**28%**)	34/99 34%	37/99 37%	12/32 (**38%**)	9/32 28%	11/32 34%
Dating matches – *unaware*	30/88 (**34%**)	24/88 27%	34/88 39%	10/26 (**38%**)	9/26 35%	7/26 27%
Dating matches – *all*	34/99 (**34%**)	28/99 28%	37/99 37%	12/31 (**39%**)	11/31 35%	8/31 26%
Emotional contagion – *unaware*	30/76 (**39%**)	20/76 26%	26/76 34%	14/24 (**58%**)	6/24 25%	4/24 17%
Emotional contagion – *all*	46/99 (**46%**)	24/99 24%	29/99 29%	19/32 (**59%**)	8/32 25%	5/32 16%
Censorship-detection webpage	51/99 (**52%**)	34/99 34%	14/99 14%	16/32 (**50%**)	12/32 38%	4/32 12%
Censorship-detection scan	53/99 (**54%**)	32/99 32%	14/99 14%	16/32 (**50%**)	10/32 31%	6/32 19%
Urinal observation	62/99 (**63%**)	21/99 21%	16/99 16%	22/32 (**69%**)	6/32 19%	3/32 9%

(b) **Disapproval of the experiment.** Q2: "Do you believe the researchers should be allowed to proceed with this experiment?"

Table 2: For each of the experimental scenarios we described to respondents, we asked two questions to gauge their concern for participants and disapproval of the experiment. As experiments are written from the perspective of researchers, we asked these questions in the order shown so as to give respondents the opportunity to think about the perspective of participants before considering whether the experiment should be allowed to proceed. We use boldface to highlight the percent who responded "no" to these questions as this answer indicates concern for participants in Q1 (a) and disapproval of the experiment in Q2 (b). Rows marked "unaware" exclude participants who were aware of the real-world experiment on which the scenario was based.

results to measure the impact of latency, were on the benign side.

Censorship-measurement experiments

Outside of our baseline concerning scenario, the two experiments that our respondents objected to the most were the censorship-detection studies from the networking measurement community. We share these results with some apprehension, giving the shaming and hostility that has been directed at researchers of other controversial studies. The researchers who conducted these studies did not, to our knowledge, circumvent any ethical regulations or requirements.

The researchers submitted their work to the Internet Measurement Conference in confidence, and upon receiving their rejection notices they could have prevented us from ever finding out about their experiments. Instead, they volunteered to identify themselves as the authors of work deemed ethically concerning and helped us to create scenarios based on their experiments.

The level of disapproval could be the result of our failure to fully convey the social benefit of the actual experiments.

Regardless of whether the disapproval reported for the censorship-detection scenarios is truly reflective of the actual experiments, they appear to present evidence of a sizeable gap in the regulatory infrastructure that is supposed to protect the public from potentially-harmful experiments. The common rule applies to "research involving human subjects" (46.101) [12], not all research that might harm humans. The scenarios presented – whether reflective of the actual experiments or not – indicate discomfort for research that can be conducted outside the jurisdiction of the regulatory regime imposed by the common rule.

The gap in our regulatory regimes is not one that can be filled by conference program committees alone; any harm resulting from an experiment is likely to occur during an experiment, and not after the results have been submitted for publication. Rather, in the long run, we need to fix the gap in our regulatory regime so that *any* research with potential to cause harm to humans is reviewed by those without a direct interest in the conduct or success of the experiment.

Until the gaps in our regulatory regimes are acknowledged and repaired, research communities may need to implement measures that can address the immediate needs of their communities. To their credit, networking researchers – including both those on the committee that rejected recent papers and the authors of those papers – are working together to address the regulatory gap for their research community, with a one-hour panel to introduce the problem at the ACM Internet Measurement Conference in November 2014 and a one-day workshop planned for ACM SigComm in July 2015.

5   Limitations

Our survey has a number of additional limitations that are important to consider when examining our results.

While we designed our survey instrument to anticipate the risks and concerns of future experimental participants, all the experiments we examined in this use of the instrument took place in the past.

As with any compression process, some fidelity will inevitably be lost when complex experiments are simplified for presentation. In describing experiments, we may have failed to anticipate which facts would be most salient to respondents. We may also have incorrectly interpreted information about an experimental design. As two authors were researchers on two of the studies described, we may have been subject to subconscious biases.

Even if survey respondents closely resemble those who would be participants in research scenarios, there's no way to guarantee that respondents will correctly anticipate how they would feel about the experiment were they to be a participant.

Finally, respondents were not required to have any prior background in ethics, ethics training, or knowledge of laws and regulations that govern research ethics (e.g., the common rule); nor did we provide them with any such background or training. This was by design. Ethical controversies can occur when there is a disconnect between what ethics boards will approve and what members of the public consider acceptable.

References

[1] Sam Burnett and Nick Feamster. Encore: Lightweight measurement of web censorship with cross-origin requests. *arXiv*, (1410.1211v1), October 5 2014.

[2] Sam Burnett and Nick Feamster. Encore: Measure web censorship (faq page). `https://encore.noise.gatech.edu/faq.html`, October 31 2014.

[3] Melanie J Clark. Re: Irb/ethics questions. Email to Nicholas G Feamster, archived by Nicholas Feamster at `https://encore.noise.gatech.edu/irb-mail.txt`, September 17 2014.

[4] Sauvik Das, Adam Kramer, Laura Dabbish, and Jason Hong. Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *The 21st ACM Conference on Computer and Communications Security (CCS)*, November 2014.

[5] Roya Ensafi, Philipp Winter, Abdullah Mueen, and Jedidiah R. Crandall. Large-scale spatiotemporal characterization of inconsistencies in the world's largest firewall. *arXiv*, (arXiv:1410.0735v1), October 2014.

[6] Nicholas Feamster. Email to stuart schechter, October 31 2014.

[7] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. Online controlled experiments at large scale. In *The 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, August 2013.

[8] R. Dennis Middlemist, Eric S. Knowles, and Charles F. Matter. Personal space invasions in the lavatory: Suggestive evidence for arousal. *Journal of Personality and Social Psychology*, 33(5):541–546), 1976.

[9] Gregg R. Murray, Cynthia R. Rugeley, Dona-Gene Mitchell, and Jeffery J. Mondak. Convenient yet not a convenience sample: Jury pools as experimental subject pools. *Social Science Research*, 42:246–253, January 2013.

[10] Christian Rudder. We experiment on human beings! `http://blog.okcupid.com/index.php/we-experiment-on-human-beings/`, July 28 2014.

[11] Stuart Schechter and Cristian Bravo-Lillo. Using ethical-response surveys to identify sources of disapproval and concern with facebook's emotional contagion experiment and other controversial studies. Technical Report MSR-TR-2014-97, July 2014.

[12] U.S. Department of Health & Human Services. Federal policy for the protection of human subjects ('common rule'). Code of Federal Regulations Title 45 Part 46 `http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html`.

A Censorship-detection webpage Experiment IRB Response

This appendix was copied from text file [3] made available by the researchers who conducted the experiment.

Request

```
From: Feamster, Nicholas G
Sent: Monday, September 15, 2014 11:35 AM
To: Clark, Melanie J
Cc: Burnett, Samuel
Subject: IRB/Ethics questions

Hi Melanie,

I would like to talk to you for 30 minutes about this project:
http://encore.noise.gatech.edu/

This is not a human-subjects experiment per se, but certain questions have come
up, and we'd like to get an authoritative answer from you/IRB about them.

Can I pay you a visit for 30 minutes this week?

Thanks,
-Nick
```

Response

After meeting , Professor Feamster received the following confirmation that IRB review was not required.

```
From: "Clark, Melanie J" <Melanie.Clark@gtrc.gatech.edu>
Subject: RE: IRB/Ethics questions
Date: September 17, 2014 at 9:00:18 AM EDT
To: "Feamster, Nicholas G" <feamster@cc.gatech.edu>

Hi Nick,

Thanks for talking about the project. As we discussed this does not require IRB
review since it doesn't meet the definition of a human subject. Please delete
the reference to the IRB in the Encore information page.

Thanks again

Melanie J. Clark, CIP
Associate Director
Office of Research Integrity Assurance
Georgia Institute of Technology
404/894-6942
http://researchintegrity.gatech.edu
```

> **Important Note**
> **The researchers have no affiliation with the King County court and this survey is in no way endorsed by the court. We strongly request that you do not let this survey distract from, or otherwise interfere with, your duties as a juror.** Please fill out the survey only during your free time, and not when you are obligated to direct your attention to your courthouse matters. Please mail the survey only during time when you are not obligated to be in the courthouse.

Survey on the Ethics of Scientific Experimentation

This survey is part of a research project being conducted by the The Ethical Research Project (www.ethicalresearch.org), a collaboration between Microsoft Research and university partners, to investigate public perceptions of scientific research practices.

If you choose to complete this survey, you will find inside descriptions of twelve hypothetical scientific experiments, which may or may not mirror real experiments that have been performed in the past. We ask that you:

- *carefully* read a short description of each experiment (350 words or fewer),
- answer four multiple-choice questions about each experiment, and
- optionally provide short explanations of your answers.

The final page of the survey contains seven demographic questions. All personal information is optional, as are explanations of your answers. While we will not attempt to identify you or associate you with your responses, we reserve the right to copy or quote the responses you provide in publications and to share your responses with other researchers.

No scientific expertise is required. The entire survey should take between 25-45 minutes of your time, depending on how many comments you provide. The $5 gratuity provided by the researcher is yours to keep even if you choose not to complete the survey. While we would be very grateful should you decide to complete the survey, you are under absolutely no obligation to do so.

If you choose to complete this survey, please send your completed survey back to us by mail in the postage-paid envelope we provided it in. If you have any questions or concerns, please contact the principal investigator for this survey, Stuart Schechter of Microsoft Research, at Stuart.Schechter@microsoft.com or at 425-*ELIDED FOR PUBLICATION*.

University researchers want to know whether choice of favorite color is influenced by one's gender.

To test this hypothesis, the researchers will perform a survey of respondents on Amazon's Mechanical Turk, a service that recruits workers for tasks that require human intelligence.

- The researchers will ask participants for their favorite color and gender.
- The researchers will ask participants to consent to be part of a study, and will not ask for any information that could reveal the identity of participants.
- The researchers will pay participants $1.00 for completing this two-question survey.
- The researchers will use statistical tests to determine whether there is a correlation.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by withholding any information that might identify them.

If the researchers are not allowed to perform this experiment, they will be unable to determine whether the commonly-assumed relationship between gender and favorite color is real or a myth.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
        Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
        Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
        Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
        Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Researchers at a search engine, such as Google web search or Microsoft's Bing, want to measure the value of speeding up their responses to users' web search queries. Reducing search response times (increasing speed) may require investments in additional computers and facilities, which may have financial and environmental costs.

Conducting such research is challenging because one cannot directly measure the benefit of an investment without first making the investment.

- Instead, the researchers will add a small delay (one to two tenths of a second) to all searches made by 10% of users, selected at random from all users.
- The researchers will use existing approaches to measure increases in users' efficiency (for example, the fraction of completed searches) and satisfaction (for example, the fraction of users who continue to use the search engine).
- The researchers will compare measurements taken from those users whose searches were delayed with measurements from users whose searches received no delay (and thus had a virtual speed improvement of those whose searches were delayed).
- Users will not be notified that they are part of the experiment nor have a chance to opt out of the experiment.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to perform this experiment, they may waste financial and environmental resources to provide speed improvements that are too small for users to notice, or they may fail to make investments in performance that their users would consider worthwhile.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
       Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
       Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
       Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
       Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Phishing is an attack in which users are sent emails with a link to a fraudulent website in order to trick them into divulging their passwords. For example, some phishing emails appear to come from a user's bank and contain a link to a website that also appears to be the user's bank, but is actually controlled by the attacker. When the user types the password into the fake site, the attacker takes the password and can now login to the user's account.

University researchers want to quantify how much the success of a phishing attack would increase if the email appeared to come from someone the recipient trusted—a friend:

- The researchers will send phishing emails to students with a link to a website that impersonates one of the university's websites.
- The researchers will send half of the students an email that appears to be from one of the student's friends, who the researchers will identify by examining the student's Facebook profile. The researchers will send the other half of students an email that appears to be sent by someone the student does not know.
- If students enter passwords into the researchers' site, the researchers will, with the permission of the university, use the university's systems to verify that the passwords entered were valid passwords.
- Afterwards, the researchers will notify students that this was a research study. They will offer students the opportunity to ask to have their data excluded from the study and to comment about the study on a blog.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to perform this experiment, they will not be able to measure how often users fall victim to phishing attacks. Therefore, the researchers will not be able to publish recommendations to help users better learn to recognize such attacks.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
    Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
    Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
    Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
    Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

   University researchers want to study whether men's urination (peeing) is affected by proximity to other men. Specifically, they want to know if being in close proximity to other men delays urination or changes to the length of time required to urinate.

   To test this hypothesis, the researchers will study a public men's restroom at their university and control for the proximity (distance) of other men using urinals. The researchers place a periscopic mirror in the floor of a toilet stall next to a urinal so that one of the researchers, within the stall, can view the lower torso and urine stream of those using the urinal. They will be unable to see the face or any other part of the body that would identify the individual using the urinal. When men enter to use the restroom, the researchers will randomly assign them to one of three groups.

- For the close-distance group, a researcher pretending to be another restroom user (a *confederate*) will use an adjacent urinal.
- For the medium-distance group, a confederate will use a urinal that is not adjacent, but one urinal away from the adjacent urinal.
- For the control group, no confederate will appear and all other urinals will be empty.
- The researchers will compare these groups by examining the delay before urination and the length of urination.
- The individuals observed by the researchers will not be notified that they are part of the experiment nor have a chance to opt out of the experiment.
- The researchers will publish the aggregate results of the experiment in a scientific paper.

   If the researchers are not allowed to perform this experiment, they may be unable to understand how intrusions on personal space affect physiological functions.

  The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
      Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
      Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
      Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
      Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Researchers at an online dating site want to determine how accurately their algorithms predict whether two people are likely to be interested in communicating. The researchers have found that pairs who the researchers predict will want to communicate do indeed end up communicating.

The challenge in measuring the accuracy of the prediction algorithm is that users see the prediction scores. The predictions may appear to work because users may be more likely to communicate with those who they are told they should like.

- To determine whether the prediction algorithm is correctly identifying pairs who want to communicate, or whether pairs communicate because the prediction algorithm suggests they should, the researchers will sometimes show users a false prediction.
- For some users predicted to be good matches, the researchers will show scores falsely indicating that they are bad matches.
- For some users predicted to be bad matches, the researchers will show scores falsely indicating that they are good matches.
- The researchers will then be able to determine how much the likelihood that two people communicate is the result of the score that their dating website displays and how much the likelihood is actually being predicted by the true score.
- After the experiment, the researchers will notify those users who were unwitting participants in the experiment to inform them of what happened. (However, participants will not be asked if they would like data collected from them to be removed from the experiment's results.)
- The researchers will publish the aggregate results of the experiment in a blog post, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to test their hypothesis, they may never know whether their users actually benefit from seeing these scores.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
    Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
    Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
    Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
    Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Computer security researchers want to learn the fraction of Internet users who fall for tricks used by hackers to steal users' passwords.

Conducting such research is challenging because if research participants know the attack is coming, or even that the study is about computer security, they may be less likely to fall for the tricks. The researchers thus plan to deceive participants as to the purpose of the human intelligence task (HIT) they will be asked to complete:

- During the task the researchers will replicate the techniques that hackers use to trick users into typing their passwords.
- Unlike criminal hackers, the researchers will not actually steal, collect, or store the passwords that users type.
- Afterwards, the researchers will present a detailed explanation of the deception to participants, reveal the true purpose of the study, and reassure participants that no passwords were actually stolen during the study.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to perform this experiment, they will not be able to measure how often users fall victim to attacks that target users' passwords. Therefore, the researchers will not be able to produce or publish recommendations that help users better learn to recognize such attacks.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
  Yes I have no preference No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
  Yes Yes, but with caution No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
  Yes No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
  Yes No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

China's government monitors citizens' Internet communications and blocks access to forbidden content. A service called Tor, funded in part by the US government, helps users in China (and throughout the world) access the Internet free from censorship and makes it harder for governments to monitor them. China tries to block citizens' access to Tor. Researchers at a US university want to measure how successful China is in blocking Tor.

Conducting such research is challenging because US researchers do not have regular access to computers throughout China.

- Instead, the researchers will identify a random set of computers located in China, which they do not own, to measure connectivity from—without consent of those computer's owners.
- The researchers will send these Chinese computers messages that trick them into replying with messages to the Tor service. These short messages contain only information used to set up communications, but no *content* meaningful to humans. No forbidden content will be sent or received.
- The researchers will use the same trick to test whether computers from the Tor service can send messages to the Chinese computers—again without causing any forbidden content to be sent.
- The researchers will publish the aggregate results of the experiment in a scientific paper and include the set of addresses of both the Tor servers and the Chinese computers

While it is possible that China's government could identify and punish those users whose computers the researchers tricked into sending messages to the Tor, there are no known instances of China prosecuting or punishing computer users for communicating with the Tor service. Furthermore, if China's monitors inspected the communications closely, they would discover that the messages contained no forbidden content and were not initiated by the computers' owners. The researchers reason that, by publishing the addresses of the Chinese computers they contacted, they will provide the Chinese users further evidence that these users did not instruct their computers to communicate with the Tor service.

If the researchers are not allowed to perform this experiment, they may be less able to help Chinese computer users access the Internet free from censorship and monitoring.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?

    Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?

    Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?

    Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?

    Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Researchers at a social network want to study whether users are more likely to share positive (happy) thoughts if their friends have been posting positive thoughts , and whether they are more likely to share negative (unhappy) thoughts if their friends have been sharing negative thoughts.

- To increase the proportion of positive posts in some users' news feeds, the researchers will randomly exclude some fraction of friends' negative posts each time the news feed is loaded.
- To increase the proportion of negative posts in some users' news feeds, the researchers will randomly exclude some fraction of friends' positive posts each time the news feed is loaded.
- The researchers will use an automated algorithm to measure whether users' posts are of a positive or negative mood.
- Users will not be notified that they are part of the experiment nor have a chance to opt out of the experiment.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to perform this experiment, they will not be able to make a valid scientific determination of whether users' moods are affected by the moods of their friends' posts. Therefore, the researchers will not be able to produce features that might protect the moods of psychologically-vulnerable users.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
    Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
    Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
    Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
    Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Computer security researchers want to measure different techniques for presenting security warnings.

One challenge in studying security decision making is that if participants are made aware that researchers are studying their security behavior, or become aware of it, they are likely to behave differently than they normally would. The researchers thus plan to deceive participants as to the purpose of the human intelligence task (HIT) they will be asked to complete:

- The researchers will give participants a task unrelated to security, but that will cause participants to encounter a simulated security warning.
- While the warning will create the illusion that the participant is facing a security risk, the researchers will not actually expose participants to any real security risks.
- The researchers will measure how different ways of presenting a warning may make that warning more or less effective in convincing users to avoid a risk.
- At the conclusion of the experiment, the researchers will present a detailed explanation of the deception to participants, reveal the true purpose of the study, and reassure participants that they were never at any real risk.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to perform this experiment, they will not be able to measure the effectiveness of different designs for computer security warnings. Therefore, the researchers will not be able to produce or publish recommendations to improve the effectiveness of future security warnings.

   The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
       Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
       Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
       Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
       Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Computer security researchers, seeking to understand the economic infrastructure that enables email spam, want to measure the rate at which spam emails result in purchases.

Conducting such research is challenging. Researchers would not want to send spam. Spammers are unlikely to divulge how successful their emails are in attracting purchases.

- The researchers will allow one of their computers to become infected with software that is controlled by spammers, while the researchers maintain sufficient control of the computer to monitor how attackers are using it.
- The researchers will alter the commands that the spammers send to the researchers' infected computer, replacing the link to the spammer's store with a link to a website run by the researchers that mimics the appearance of the spammer's store.
- Without collecting payments or other personal information about those users who respond to the spam email seeking to make a purchase from the spammers, the researchers record the number of attempts made to purchase products from the store advertised by the spam. If users try to pay for goods, the researchers will show an error message.
- The researchers will not inform those users who receive the spam sent by attackers using the infected computer as this might cause users to behave differently or otherwise compromise the validity of the results.
- The researchers will not inform those users who visit the store to make a purchase that the store has been disabled, or that their choice to make a purchase is being recorded.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by removing any information that might identify them.

If the researchers are not allowed to perform this experiment, they will not be able to empirically measure the effectiveness of spam emails and may not be able to produce or publish well-informed recommendations for technical or policy approaches to stopping spam.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
      Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
      Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
      Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
      Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Researchers at a social network want to know whether users will be more likely to use important security protections if these users learn that many of their friends already use these protections.

The researchers will test this hypothesis on 50,000 users, selected at random from those users who have at least ten friends using the security protection. They will assign users to groups using random assignment.

- The researchers will place an announcement, directly above the news feeds of users in all groups, inviting each user to improve their account security by adding a security protection.
- One group of users will receive an announcement that includes no information about how many of their friends are also using this protection.
- Other groups will receive announcements that contain information about how many of their friends are already using the protection. Some users will be told the number of friends who use the protection and others will be told the percentage of friends.
- The researchers exclude from the study those users who have fewer than ten friends using the security protection, as these users might be able to figure out which friends are using the protection and infer which friends are not.
- The researchers will compare these groups by examining the percentage who clicked on the announcement and the percentage who completed the process of adopting the security protection.
- Users will not be notified that they are part of the experiment nor have a chance to opt out of the experiment.
- The researchers will publish the aggregate results of the experiment in a scientific paper, keeping participants anonymous by withholding any information that might identify them.

If the researchers are not allowed to perform this experiment, they may not have the information they need to best encourage users to adopt security protections that can safeguard them from harm.

   The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?
        Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?
        Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?
        Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?
        Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

Some countries censor their citizens' access to websites; they monitor the address to which citizens' internet communications are directed and block addresses of sites they disapprove of. Researchers at a US university have a list of websites that they suspect countries sometimes censor. They would like to determine when countries are censoring each website.

Conducting such research is challenging because US researchers may not have regular access to computers in the countries they wish to study.

- Instead, the researchers will modify their own websites, which are sometimes visited by users in other countries.
- The researchers will record the Internet address of each visitor to their website to match it to the country they are connecting from.
- The researchers will link their web pages to content on potentially-censored websites. These links are similar to the links many websites use to include images and other content from advertisers. However, instead of instructing the users' browser to get content from advertisers, the links connect to content on the potentially-censored websites.
- The researchers will include code within their web pages that checks whether the content from potentially-censored websites was downloaded successfully, which would indicate that the website was not censored for that visitor. If none of the visitors in a country can download content from a website, that website has likely been censored.
- The researchers will publish the aggregate results of the experiment in a scientific paper, without revealing specifics that might identify individual visitors.

One risk of this study is that, when foreign users access the researchers' websites, their browser's attempt to download censored content will be recorded by government surveillance. Their government may conclude they were intentionally trying to access censored content. These visitors will not be notified that this experiment is taking place, so they will not have knowingly violated censorship restrictions.

If the researchers are not allowed to perform this experiment, they may be unable to track the ongoing use of censorship by repressive governments with as much accuracy.

The following questions concern the experiment described on the previous page. For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

If someone you cared about were a candidate participant for this experiment, would you want that person to be included as a participant?

    Yes   I have no preference   No

Please explain your answer. (optional)

Do you believe the researchers should be allowed to proceed with this experiment?

    Yes   Yes, but with caution   No

Please explain your answer. (optional)

Are you aware of having ever participated in such a study?

    Yes   No

If so, please explain approximately when and where (if in person) you participated in this study.

Are you aware of the study described, or one similar to it, having been performed by researchers in the past? For example, have you have heard about it in the news or learned about it in a class?

    Yes   No

If so, please explain how you learned about this study.

Is there anything that we could have made clearer in describing this experiment, or any additional information that would have helped you answer the questions above?

These last questions are about your age, gender, occupation, and online experiences. We realize these questions may be somewhat personal. We would greatly appreciate your answers so we can determine what factors may have influenced your responses.

For each multiple choice question, please circle exactly one answer. Please do not leave questions unanswered or circle more than one answer.

In what year were you born?

What is your gender?

    Male   Female   I'm uncomfortable answering

What is your occupation?

Have you ever purchased goods advertised via an unsolicited marketing email?

    Yes   No   I'm uncomfortable answering

Have you ever participated in a study that involved deception?

    Yes   No   I'm uncomfortable answering

If you have participated in studies that involved deception, please explain how many and describe one.

Please indicate whether your annual household income is below or above $53,000. (You are not required to answer this question but, if you can, it will be of great value in helping us determine if we are gathered opinions from an economically-diverse sample.)

    Below $53,000   At or above $53,000   I'm uncomfortable answering

Prior to receiving this survey, had you heard about Facebook's emotional contagion study, sometimes known as the Facebook 'mood' study? (This is the experiment that was the subject of many news stories in the summer of 2014.)

    Yes   No

Prior to participating in this study, had you heard about the experiment in which a dating website (OKCupid) changed the match scores shown to users?

    Yes   No

Thank you for completing the survey.

Please return it to us by US Mail in the postage-paid envelope in which we provided the survey. You may mail it to us from *any* US Mailbox, including:

- the one on the second floor of the courthouse,
- the outdoor blue mail box at the corner of Columbia and 4th Avenue,
- your home mailbox, or
- your local post office.

Again, please do not let mailing this distract you from your juror duties and do so only when you are not required to be in the courthouse, such as during breaks, lunch time, or at the end of the day.

If you are willing to write the date and sign your name below, you will make it easier for our researchers to obtain reimbursement for the $5 gratuity we provided you with. As this is an anonymous survey, you are under no obligation to provide your signature.