Using Ethical-Response Surveys to Identify Sources of Disapproval and Concern with Facebook's Emotional Contagion Experiment and Other Controversial Studies

Highly-preliminary Working paper

(expect frequent and significant changes)

Stuart Schechter
Microsoft Research

Cristian Bravo-Lillo
Carnegie Mellon University

July 10, 2014

Abstract

We surveyed 3570 workers on Amazon's Mechanical Turk to gauge their ethical response to five scenarios describing scientific experiments—including one scenario describing Facebook's emotional contagion experiment. We will post an update of this paper containing the results and analysis on or after 12:01AM Pacific on Monday July 14.

1 Introduction

In evaluating the ethicality of an experiment, researchers and ethics boards must weigh the benefits of the study against potential risks—many of which are borne by participants. Alas, there is a great deal of guesswork in anticipating how participants and others will react to an experiment. The information researchers and ethics boards need to make sound judgements that are hard to come by; researchers rarely share, or even measure, participants' feelings, concerns, and opinions of the ethicality of the experiments they take part in. The rare instances in which we learn about the ethical consequences of experiments typically occur when concerns or harms are so serious as to come to the attention of the public.

In 2012, we began using surveys to identify disapproval and concern with experiments *before* exposing participants to them. We presented respondents with a series of short descriptions of experimental scenarios and asked questions to gauge their ethical response. We wrote these short summaries with goal of packaging the information salient to evaluating the ethicality of a study into a small package understandable to a general audience. Our first such survey caused us to re-evaluate how participants might react to a study we had planned (and received approval to conduct). In light of our survey data, we concluded that the benefits to society of running the experiment no longer appeared to outweigh the risks. We first publicly advocated the prophylactic use of ethical-response surveys in 2013 [3].

In this new work, we ask what researchers at Facebook would have learned had they had the opportunity to use our ethical-response survey design prior to performing their 2012 emotional contagion experiment [6]. In Facebook's experiment, researchers used an algorithm to remove posts from users' news feeds in order to determine whether a reduction in positive or negative posts from participants' friends would impact the emotional mood of posts made by participants themselves. This experiment, which was published in June 2014, quickly became controversial—attracting criticism that the researchers and those overseeing their work presumably had not anticipated.

We performed an ethical-response survey on a convenience sample of 3570 workers on Amazon's Mechanical Turk who were based in the United States from July 2–4, 2014. Of these, 2127 reported not yet being aware of Facebook's emotional contagion experiment; these participants presented us with an opportunity to gauge opinions not yet tainted by media coverage and the evolving public reaction that has followed the publication of Facebook's experiment.

We presented participants with five experimental scenarios: one about the Facebook experiment and four about other experiments. The details of the scenario related to the Facebook experiment varied between respondents whereas the other four did not. We wrote our control scenario to describe the experiment based on our understanding of Facebook's experiment from reading their paper [6], whereas for other treatments we modified certain facts about the experiment; we modified facts such as what manipulations Facebook's researchers had performed or even which company had performed the research.

Of the other four experimental scenarios, two described deception experiments that members of our team had led in the past and for which we'd worked to measure participants' ethical response after debriefing. The final two scenarios summarized research from the past decade that had been the subject of ethical debate within the research community. All four were conducted with approval from university ethics boards.

For each of the five abstracts we presented to each respondent, we asked two questions designed to gauge concern for participants and disapproval with allowing the study. The participant *concern* question asked whether the respondent would want someone they cared about to be included as a participant. The *disapproval* question asked whether the respondent believed the experiment should be allowed to proceed or not.

2 Experimental procedure

After piloting on July 1, we offered the final draft of our survey for three days starting at 12:00AM the morning of Wednesday July 2, EDT. We used a single Human Intelligence Task on Amazon's Mechanical Turk to prevent the same worker account from taking the survey twice (though we cannot guarantee some workers with multiple accounts did not do so). We restricted workers to those coming from the United States.

After brief instructions, we presented five experimental scenarios in random order (randomized for each participant). Four of the scenarios were the same for each respondent, but we randomly assigned each respondent one of ten variants of the Facebook experiment. We then asked follow-up questions.

2.1 Recruiting

We offered a Human Interactive Task (HIT) on Mechanical Turk which presented prospective respondents with the following offer[1]:

> In this survey, you will be ask you about five hypothetical scientific experiments. For each, we will ask you to:
>
> - *Carefully* read an abstract description of the experiment (350 words or fewer).
> - Answer 4 multiple-choice questions about each experiment.
> - Optionally provide short explanations of your answers.
>
> Finally, we will ask you some brief demographic questions at the end of the study. All personal information (e.g., age) is optional. Your responses will be kept anonymous, though we reserve the right to copy or quote the responses you provide.
> The entire survey should take **under 10 minutes of your time** and pays **$1.00**.
> This survey is part of a research project being conducted by the The Ethical Research Project.
> If you have any questions, please feel free to contact us at: team@ethicalresearch.org

2.2 Instructions

Participants who accepted the HIT received the following instructions:

> Each of the following five pages will contain a description of a hypothetical scientific experiment, followed by questions about that experiment. In order to answer the questions, please read the description of each experiment carefully.

2.3 Questions for each scenario

While we randomized the order of the experimental abstracts, we kept the ordering of questions and response options consistent.

The first question that followed each scenario we designed to measure 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 designed these options to be ordinal: from least concerned to most concerned. We asked this concern question first in hopes that it would give respondents a chance to humanize potential participants and think about the consequences of the experiment on them.

We designed the second question to gauge whether respondents would disapprove of the experiment. We asked, "Do you believe the researchers should be allowed to proceed with this experiment?" We offered four options, again ordered from most approving to least approving with the first option being "Yes" and the last "No". We included the second option, "Yes,

[1] The error "you will be ask you" in place of "we will ask you" is in the original and not a transcription error. Fortunately, we described the task again in the first page of the survey.

but with caution", for respondents who did not want to disapprove of a experiment but feared that an unambiguous "yes" would relieve researchers to their duty to take their ethical duties seriously. The option between the two "Yes" options and "No" was "I'm not sure." We treat this an ordinal value between the yes and the no options as the respondent is unable to commit to either and is therefore likely to be somewhere in between.

For each of the first two questions, we gave respondents a free-response field in which to 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 by researchers in the past? (For example, have you have heard about it in the news or learned about it in a class?)"

2.4 Closing questions

After collecting respondents' responses to the five experimental scenarios, we asked the following questions about respondents' demographics and about factors that might impact their opinions:

- What year were you born?
  (please use a four-digit year, or 'd' if you decline to answer)
- 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}
- Prior to participating in this study, had you heard about Facebook's 'mood' study (the experiment that has the subject in many recent news stories).
  {Yes; No}

We placed the question about prior knowledge of Facebook's experiment at the very end of our survey so as to avoid having this question taint responses to earlier questions.

3 Experimental scenarios

We created two scenarios for experiments from the past decade that were the subject of ethical debate in the research community, two scenarios for experiments that we had run and gauged participants' ethical response to at the time of the experiment, and one scenario for the recent Facebook experiment. In no description of these experimental scenarios did we mention that the experiment described was a real experiment or that, in the case of the university studies, it had been approved by an ethics board.

A Social phishing

We wrote this experimental scenario around the "Social Phishing" experiment performed by researchers at Indiana University [4]. In their experiment, researchers sent students phishing emails to see if they could be deceived into revealing their passwords on a website that impersonated a university system. Some of the emails researchers sent were customized based on participants' public Facebook profiles. The researchers collected passwords from those who entered them and tested them against a university password database to determine if they were valid. The exact wording of this scenario is in Appendix A.A.

We did not mention that participants were exposed to the experiment without their consent.

B Spam infrastructure infiltration & analysis

This experimental scenario describes an experiment to measure the economics of spam performed by researchers at the University of California [5]. In this experiment, the researchers allowed a computer to be infected with software used to send spam. The researchers then modified the spam to direct recipients to servers controlled by the researchers, instead of the spammers. Thus, recipients of attackers' spam became unwitting participants in this study. The exact wording of this scenario is in Appendix A.B.

As with the previous study, we did not explicitly state that spam recipients did not opt into the study via a consent form, though we did indicate that spam recipients who visited the impersonated store would not be informed that it was not the genuine store run by spammers.

C Password-dialog spoofing

This scenario describes an experiment by researchers at Carnegie Mellon University and Microsoft Research to determine whether malicious websites can trick users into revealing their device (computer) password by mimicking (spoofing) security dialogs that are normally generated by the device's operating system [2]. The researchers presented the experiment to participants as an evaluation of online gaming websites. When participants visited a website run by the researchers, the researchers mimicked the operating system window used to download a software component. The window indicated that it required the user's (participant's) device username and password to install the software component. The researchers observed whether participants could be deceived to enter that information. (Unlike the Indiana University phishing study, the researchers did not actually collect passwords without participants consent.) The exact wording of this scenario is in Appendix A.C.

The experiment on which this scenario was run by a team that includes two authors of our ethical-response survey (and the paper you are reading now). The experiment was was approved by the Institutional Review Board of Carnegie Mellon University.

Participants in the actual experiment had received a consent form explaining that they were part of a University experiment, though the consent form did not disclose that security was the focus of the experiment. The researchers informed study participants of the deception during a debriefing at the end of the experiment. We elided the presence of the consent form in order to make the scenario more similar to the other, more controversial, experiments described in this survey.

D Spoofed-warning deception

This scenario describes an experiment by researchers at Carnegie Mellon University and Microsoft Research to improve security warning dialogs [1]. Like the previous study, it is a deception experiment in which researchers led participants to believe that online games were the focus of the study. Unlike the previous study, users were not tricked into typing passwords. Rather, they were shown a warning about the risk of installing software and the researchers tested to see whether participants could identify signs of danger in the warning. Regardless of how participants responded to the install warning, no harm would come to them. The exact wording presented of the scenario is in Appendix A.D.

As with the previous scenario, the experiment on which this scenario was run by a team that includes two authors of our ethical-response survey (and the paper you are reading now). The experiment was was approved by the Institutional Review Board of Carnegie Mellon University. Participants in the actual experiment had received a consent form explaining that they were part of a university experiment, though the consent form did not disclose that security was the focus of the experiment. Further, the researchers collected data to monitor participants' ethical responses during the study to ensure harm was minimal. We elided these facts in order to make the scenario more similar to the more controversial experiments described in this survey.

Researchers at Facebook 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.
- The researchers will publish the anonymized aggregate results of the experiment in a scientific paper.
- Participants will not be identified and will remain anonymous.

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.

Figure 1: Experimental scenario for Facebook's emotional contagion experiment

F Facebook's emotional contagion experiment

This scenario, presented in Figure 1, describes Facebook's emotional contagion experiment, based on our understanding of the experiment from reading their paper. The scenario focuses on facts about the experimental goals and methodology and so avoids touching on many issues that have been a subject of public debate. Specifically, it does not discuss oversight, terms of service, or the participation of university researchers in the experiment. As is consistent with the other scenarios, we do not explicitly state that the researchers did not obtain consent from participants.

However, many respondents did not receive this exact scenario (our control), but instead received one of the variants (treatments) that are described in the next section.

4 Treatments

We created ten variants of the experimental scenario for the Facebook experiment. We assigned respondents to scenario variants (treatments) at random with uniform probabilities assigned to each.

F0 Control

The control does not diverge from the facts of Facebook's experiment as we understood them, described in Section 3.F and detailed in Figure 1.

F1 Only remove positive posts

We designed this scenario to test the hypothesis that respondents might be particularly concerned with removing negative posts. We wondered if respondents would be particularly concerned that participants might miss out on important bad news (e.g., the death of a friend, or a post by a distressed friend in need of support) but less concerned about missing good news. We thus removed references to removing *negative* posts for the purpose of increasing the proportion of *positive* posts in the feed. From the first paragraph, we removed the string: "are more likely to share positive (happy) thoughts if their friends have been posting positive thoughts, and whether they". We also removed the first bullet point, which had stated: "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."

F2 Only remove negative posts

We designed this scenario to test the hypothesis that respondents might be particularly concerned with participants missing out on good news. In this treatment, participants would only miss out on negative posts. We removed references to removing *positive* posts for the purpose of increasing the proportion of *negative* posts in the feed. From the first paragraph, we removed the string: "and whether they are more likely to share negative (unhappy) thoughts if their friends have been sharing negative thoughts". We also removed the second bullet point, which had stated: "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."

F3 Remove mention of publication

We created this scenario to test whether respondents would feel more or less favorably if the mention of a scientific publication were removed. Specifically, we removed the second-to-last bullet point of the scenario, which had stated "The researchers will publish the anonymized aggregate results of the experiment in a scientific paper."

F4 Remove mention of product improvement

We created this scenario to test whether respondents would feel less favorably about the experiment if there were no mention of potential for product improvement that might benefit users. We removed the last sentence of the scenario, which had said that a consequence of not allowing the research would be that "the researchers will not be able to produce features that might protect the moods of psychologically-vulnerable users."

F5 Promise not to use for advertising

To test the hypothesis that respondents might respond more favorably to the experiment if the results would not be used for advertising, we created a scenario in which researchers promised this. We appended one item to the list of bullet points. It stated: "The researchers promise in writing that the research findings will be used only to further science and improve the product for users. The results will not be used to improve Facebook's advertising algorithms."

F6 Insert posts instead of hiding

We hypothesized that respondents might be less concerned about researchers manipulating news feeds if they only add extra (bonus) posts, as opposed to removing that had been deemed relevant to them by Facebook's regular algorithms. We changed the description of the study design so that, instead of hiding posts, the researchers would add negative or positive posts that otherwise would not have been deemed worthy of display on the news feed. We rewrote the first two bullet points as follows:

- To increase the proportion of positive posts in some users' news feeds, the researchers will randomly include additional positive posts that would otherwise have been deemed insufficiently relevant or unimportant.
- To increase the proportion of negative posts in some users' news feeds, the researchers will randomly include additional negative posts that would otherwise have been deemed insufficiently relevant or unimportant.

F7 Insert posts, and only positive ones

We hypothesized that respondents might be even less concerned if the added posts were only positive posts. We started with the prior treatment (F6), and removed from the first paragraph the string: "and whether they are more likely to share negative (unhappy) thoughts if their friends have been sharing negative thoughts". We kept the first bullet point from the prior treatment (F6), but removed the second.

F8 Replace 'Facebook' with 'a social network'

To test whether respondents' opinions would change if the experiment were not identified as being conducted by Facebook, we replaced the third word of the scenario, "Facebook", with the phrase "a social network".

F9 Replace 'Facebook' with 'Twitter'

To test whether respondents might be have responded differently to the experimental scenario had it been conducted by Twitter, we replaced the third word of this scenario, "Facebook", with "Twitter".

5 Results

We will post an update of this paper containing the results and analysis on or after 12:01AM Pacific on Monday July 14.

References

[1] C. Bravo-Lillo, L. Cranor, J. Downs, S. Komanduri, R. Reeder, S. Schechter, and M. Sleeper. Your attention please: Designing security-decision uis to make genuine risks harder to ignore. In *Proceedings of the 9th Symposium On Usable Privacy and Security*, SOUPS 2013, New York, NY, USA, 2013. ACM.

[2] C. Bravo-Lillo, L. F. Cranor, J. Downs, S. Komanduri, S. Schechter, and M. Sleeper. Operating system framed in case of mistaken identity. In *The 19th ACM Conference on Computer and Communications Security (CCS)*, Oct. 16–18 2012.

[3] C. Bravo-Lillo, S. Egelman, C. Herley, S. Schechter, and J. Tsai. You Needn't Build That: Reusable Ethics-Compliance Infrastructure for Human Subjects Research. *Cyber-security Research Ethics Dialog & Strategy Workshop*, May 2013.

[4] T. N. Jagatic, N. a. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[5] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: an empirical analysis of spam marketing conversion. *ACM Conference on Computer and Communications Security*, pages 3–14, 2008.

[6] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Science*, 111(24):8788–8790, 2014.

A Verbatim scenarios

A Social phishing

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 its targets received appeared to come from someone the target user 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 inform 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 anonymized aggregate results of the experiment in a scientific paper.
- Participants will not be identified and will remain anonymous.

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.

B Spam infrastructure infiltration & analysis

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.
- The researchers will not inform 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 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 anonymized aggregate results of the experiment in a scientific paper.
- Participants will not be identified and will remain anonymous.

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.

C Password-dialog spoofing

Computer security researchers want to learn the fraction of Internet users who fall for the 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 anonymized aggregate results of the experiment in a scientific paper.
- Participants will not be identified and will remain anonymous.

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.

D Spoofed-warning deception

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 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 anonymized aggregate results of the experiment in a scientific paper.
- Participants will not be identified and will remain anonymous.

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.