Everything We Know About Facebook Secret Mood Manipulation Experiment

Was it ethical?

ROBINSON MEYER

REUTERS

Updated, 09/08/14

Facebook's News Feed—the main list of status updates, messages, and photos you see when you open Facebook on

your computer or phone—is not a perfect mirror of the world.

But few users expect that Facebook would change their News Feed in order to manipulate their emotional state.

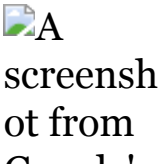
We now know that's exactly what happened two years ago. For one week in January 2012, data scientists skewed what almost 700,000 Facebook users saw when they logged into its service. Some people were shown content with a preponderance of happy and positive words; some were shown content analyzed as sadder than average. And when the week was over, these manipulated users were more likely to post either especially positive or negative words themselves.

This tinkering was just revealed as part of a new study, published in the prestigious *Proceedings of the National Academy of Sciences*. Many previous studies have used Facebook data to examine "emotional contagion," as this

MORE STORIES

Just This One Time, Tech's Not the Main Problem
ALEXIS C. MADRIGAL

Google's New Tool to Fight Climate Change
ROBINSON MEYER



With Instagram's Founders Out, Welcome to Facebook Inc.
ALEXIS C. MADRIGAL

When Televisions Were Radioactive
SUSAN MURRAY

one did. This study is different because, while other studies have observed Facebook user data, this one set out to manipulate it.

The experiment is almost certainly legal. In the company's current terms of service, Facebook users relinquish the use of their data for "data analysis, testing, [and] research." Is it ethical, though? Since news of the study first emerged, I've seen and heard both privacy advocates and casual users express surprise at the audacity of the experiment.

> In the wake of both the Snowden stuff and the Cuba twitter stuff, the Facebook "transmission of anger" experiment is terrifying.
>
> — Clay Johnson (@cjoh) June 28, 2014

> Get off Facebook. Get your family off Facebook. If you work there, quit. They're fucking awful.
>
> — Erin Kissane (@kissane) June 28, 2014

We're tracking the ethical, legal, and philosophical response to this Facebook experiment here. We've also asked the authors of the study for comment. Author Jamie Guillory replied and referred us to a Facebook spokesman. Early Sunday morning, a Facebook spokesman sent this comment in an email:

> This research was conducted for a single week in 2012 and none of the data used was associated with a specific person's Facebook account. We do research to improve our services and to make the content people see on Facebook as relevant and engaging as possible. A big part of this is understanding how people respond to different types of content, whether it's positive or negative in tone, news from friends, or information from pages they follow. We carefully consider what research we do and have a strong internal review process. There is no unnecessary collection of people's data in connection

> with these research initiatives and all data is stored securely.

And on Sunday afternoon, Adam D.I. Kramer, one of the study's authors and a Facebook employee, commented on the experiment in a public Facebook post. "And at the end of the day, the actual impact on people in the experiment was the minimal amount to statistically detect it," he writes. "Having written and designed this experiment myself, I can tell you that our goal was never to upset anyone. [...] In hindsight, the research benefits of the paper may not have justified all of this anxiety."

Kramer adds that Facebook's internal review practices have "come a long way" since 2012, when the experiment was run.

What did the paper itself find?

The study found that by manipulating the News Feeds displayed to 689,003 Facebook users users, it could affect the

content which those users posted to Facebook. More negative News Feeds led to more negative status messages, as more positive News Feeds led to positive statuses.

As far as the study was concerned, this meant that it had shown “that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness.” It touts that this emotional contagion can be achieved without “direct interaction between people” (because the unwitting subjects were only seeing each others’ News Feeds).

The researchers add that never during the experiment could they read individual users’ posts.

Two interesting things stuck out to me in the study.

The first? The effect the study documents is very small, as little as one-tenth of a percent of an observed change. That doesn’t mean it’s unimportant, though, as the authors add:

> Given the massive scale of social networks such as Facebook, even small effects can have large aggregated consequences. […] After all, an effect size of d = 0.001 at Facebook's scale is not negligible: In early 2013, this would have corresponded to **hundreds of thousands of emotion expressions in status updates per day**.

The second was this line:

> Omitting emotional content reduced the amount of words the person subsequently produced, both when positivity was reduced ($z = -4.78$, $P < 0.001$) and when negativity was reduced ($z = -7.219$, $P < 0.001$).

In other words, when researchers reduced the appearance of *either* positive or negative sentiments in people's News Feeds—when the feeds just got generally less emotional—those people stopped writing so many words on Facebook.

Make people's feeds blander and they stop typing things

into Facebook.

Was the study well designed?

Perhaps not, says John Grohol, the founder of psychology website Psych Central. Grohol believes the study's methods are hampered by the misuse of tools: Software better matched to analyze novels and essays, he says, is being applied toward the much shorter texts on social networks.

> Let's look at two hypothetical examples of why this is important. Here are two sample tweets (or status updates) that are not uncommon:
>
> "I am not happy.
>
> "I am not having a great day."
>
> An independent rater or judge would rate these two tweets as negative — they're clearly expressing a negative emotion. That would be +2 on the negative scale, and 0 on the positive scale.

> But the LIWC 2007 tool doesn't see it that way. Instead, it would rate these two tweets as scoring +2 for positive (because of the words "great" and "happy") and +2 for negative (because of the word "not" in both texts).

"What the Facebook researchers clearly show," writes Grohol, "is that they put too much faith in the tools they're using without understanding — and discussing — the tools' significant limitations."

Did an institutional review board (IRB)—an independent ethics committee that vets research that involves humans—approve the experiment?

According to a Cornell University press statement on Monday, the experiment was conducted before an IRB was consulted.* Cornell professor Jeffrey Hancock—an author of the study—began working on the results *after* Facebook had

conducted the experiment. Hancock only had access to results, says the release, so "Cornell University's Institutional Review Board concluded that he was not directly engaged in human research and that no review by the Cornell Human Research Protection Program was required."

In other words, the experiment had already been run, so its human subjects were beyond protecting. Assuming the researchers did not see users' confidential data, the results of the experiment could be examined without further endangering any subjects.

Both Cornell and Facebook have been reluctant to provide details about the process beyond their respective prepared statments. One of the study's authors told *The Atlantic* on Monday that he's been advised by the university not to speak to reporters.

By the time the study reached Susan Fiske, the Princeton University psychology professor who edited the study

for publication, Cornell's IRB members had already determined it outside of their purview.

Fiske had earlier conveyed to *The Atlantic* that the experiment was IRB-approved.

"I was concerned," Fiske told *The Atlantic* on Saturday, "until I queried the authors and they said their local institutional review board had approved it—and apparently on the grounds that Facebook apparently manipulates people's News Feeds all the time."

On Sunday, other reports raised questions about how an IRB was consulted. In a Facebook post on Sunday, study author Adam Kramer referenced only "internal review practices." And a Forbes report that day, citing an unnamed source, claimed that Facebook only used an internal review.

When *The Atlantic* asked Fiske to clarify Sunday, she said the researchers' "revision letter

said they had Cornell IRB approval as a 'pre-existing dataset' presumably from FB, who seems to have reviewed it as well in some unspecified way... Under IRB regulations, pre-existing dataset would have been approved previously and someone is just analyzing data already collected, often by someone else."

The mention of a "pre-existing dataset" here matters because, as Fiske explained in a follow-up email, "presumably the data already existed when they applied to Cornell IRB." (She also noted: "I am not second-guessing the decision.") Cornell's Monday statement confirms this presumption.

On Saturday, Fiske said that she didn't want the "the originality of the research" to be lost, but called the experiment "an open ethical question."

"It's ethically okay from the regulations perspective, but ethics are kind of social decisions. There's not an absolute answer. And so the

level of outrage that appears to be happening suggests that maybe it shouldn't have been done...I'm still thinking about it and I'm a little creeped out, too."

For more, check *Atlantic* editor Adrienne LaFrance's full interview with Prof. Fiske.

From what we know now, were the experiment's subjects able to provide *informed consent?*

In its ethical principles and code of conduct, the American Psychological Association (APA) defines *informed consent* like this:

> When psychologists conduct research or provide assessment, therapy, counseling, or consulting services in person or via electronic transmission or other forms of communication, they obtain the informed consent of the individual or individuals using language that is reasonably understandable to that person or persons except when conducting such

> activities without consent is mandated by law or governmental regulation or as otherwise provided in this Ethics Code.

As mentioned above, the research seems to have been carried out under Facebook's extensive terms of service. The company's current data use policy, which governs exactly how it may use users' data, runs to more than 9,000 words and uses the word "research" twice. But as *Forbes* writer Kashmir Hill reported Monday night, the data use policy in effect when the experiment was conducted never mentioned "research" at all—the word wasn't inserted until May 2012.

Never mind whether the current data use policy constitutes "language that is reasonably understandable": Under the January 2012 terms of service, did Facebook secure even shaky consent?

The APA has further guidelines for so-called "deceptive research" like this, where the

real purpose of the research can't be made available to participants during research. The last of these guidelines is:

> Psychologists explain any deception that is an integral feature of the design and conduct of an experiment to participants as early as is feasible, preferably at the conclusion of their participation, but no later than at the conclusion of the data collection, and permit participants to withdraw their data.

At the end of the experiment, did Facebook tell the user-subjects that their News Feeds had been altered for the sake of research? If so, the study never mentions it.

James Grimmelmann, a law professor at the University of Maryland, believes the study did not secure informed consent. And he adds that Facebook fails even its own standards, which are lower than that of the academy:

> A stronger reason is that even when Facebook

> manipulates our News Feeds to sell us things, it is supposed—legally and ethically—to meet certain minimal standards. Anything on Facebook that is actually an ad is labelled as such (even if not always clearly.) This study failed even that test, and for a particularly unappealing research goal: *We wanted to see if we could make you feel bad without you noticing. We succeeded.*

Did the U.S. government sponsor the research?

Cornell has now updated their June 10 story to say that the research received no external funding. Originally, Cornell had identified the Army Research Office, an agency within the U.S. Army that funds basic research in the military's interest, as one of the funders of their experiment.

Do these kind of News Feed tweaks happen at other times?

At any one time, Facebook said last year, there were on average 1,500 pieces of content that could show up in your News Feed.

The company uses an algorithm to determine what to display and what to hide.

It talks about this algorithm very rarely, but we know it's very powerful. Last year, the company changed News Feed to surface more news stories. Websites like BuzzFeed and Upworthy proceeded to see record-busting numbers of visitors.

So we know it happens. Consider Fiske's explanation of the research ethics here—the study was approved "on the grounds that Facebook apparently manipulates people's News Feeds all the time." And consider also that *from this study alone* Facebook knows at least one knob to tweak to get users to post more words on Facebook.

---

*

This post originally stated that an institutional review board, or IRB, was consulted before the experiment took place regarding certain aspects of data collection.

Adrienne LaFrance contributed writing and reporting.