THE CONVERSATION
Academic rigor, journalistic flair

Our experiments taught us why people troll

March 1, 2017 8.39pm EST

Trolling can spread from person to person. Cropped from Ayana T. Miller/flickr, CC BY-ND

"Fail at life. Go bomb yourself."

Comments like this one, found on a CNN article about how women perceive themselves, are prevalent today across the internet, whether it's Facebook, Reddit or a news website. Such behavior can range from profanity and name-calling to personal attacks, sexual harassment or hate speech.

A recent Pew Internet Survey found that four out of 10 people online have been harassed online, with far more having witnessed such behavior. Trolling has become so rampant that several websites have even resorted to completely removing comments.

Many believe that trolling is done by a small, vocal minority of sociopathic individuals. This belief has been reinforced not only in the media, but also in past research on trolling, which focused on interviewing these individuals. Some studies even showed that trolls have predisposing personal and biological traits, such as sadism and a propensity to seek excessive stimulation.

But what if all trolls aren't born trolls? What if they are ordinary people like you and me? In our research, we found that people can be influenced to troll others under the right circumstances in an online community. By analyzing 16 million comments made on CNN.com and conducting an online controlled experiment, we identified two key factors that can lead ordinary people to troll.

What makes a troll?

We recruited 667 participants through an online crowdsourcing platform and asked them to first take a quiz, then read an article and engage in discussion. Every participant saw the same article, but some were given a discussion that had started with comments by trolls, where others saw neutral comments instead. Here, trolling was defined using standard community guidelines – for example, name-calling, profanity, racism or harassment. The quiz given beforehand was also varied to either be easy or difficult.

Authors

Justin Cheng
Ph.D Student in Computer Science, Stanford University

Cristian Danescu-Niculescu-Mizil
Assistant Professor of Information Science, Cornell University

Jure Leskovec
Associate Professor of Computer Science, Stanford University

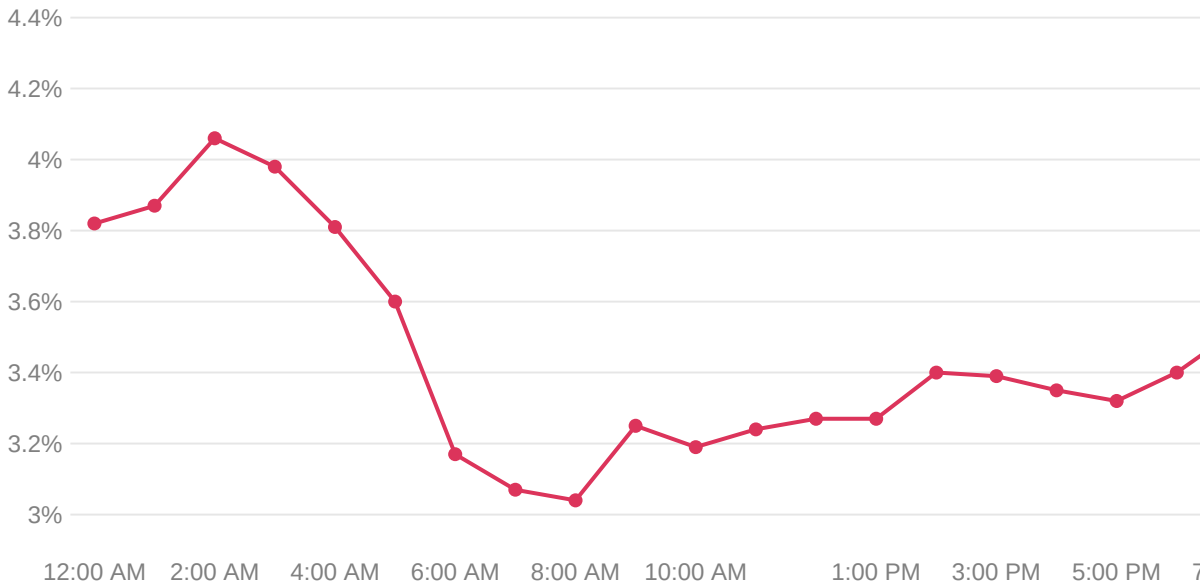
Michael Bernstein
Assistant Professor of Computer Science, Stanford University

Our analysis of comments on CNN.com helped to verify and extend these experimental observations.

The first factor that seems to influence trolling is a person's mood. In our experiment, people put into negative moods were much more likely to start trolling. We also discovered that trolling ebbs and flows with the time of day and day of week, in sync with natural human mood patterns. Trolling is most frequent late at night, and least frequent in the morning. Trolling also peaks on Monday, at the beginning of the work week.

Time to troll

The proportion of flagged posts peaks late at night, according to a study of comments on CNN.com.



The Conversation, CC-BY-ND

Source: Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions (CSCW 2017) Get the data

Moreover, we discovered that a negative mood can persist beyond the events that brought about those feelings. Suppose that a person participates in a discussion where other people wrote troll comments.

If that person goes on to participate in an unrelated discussion, they are more likely to troll in that discussion too.

The second factor is the context of a discussion. If a discussion begins with a “troll comment,” then it is twice as likely to be trolled by other participants later on, compared to a discussion that does not start with a troll comment.

In fact, these troll comments can add up. The more troll comments in a discussion, the more likely that future participants will also troll the discussion. Altogether, these results show how the initial comments in a discussion set a strong, lasting precedent for later trolling.

We wondered if, by using these two factors, we could predict when trolling would occur. Using machine learning algorithms, we were able to forecast whether a person was going to troll about 80 percent of the time.

Interestingly, mood and discussion context were together a much stronger indicator of trolling than identifying specific individuals as trolls. In other words, trolling is caused more by the person’s environment than any inherent trait.

Since trolling is situational, and ordinary people can be influenced to troll, such behavior can end up spreading from person to person. A single troll comment in a discussion – perhaps written by a person who woke up on the wrong side of the bed – can lead to worse moods among other participants, and even more troll comments elsewhere. As this negative behavior continues to propagate, trolling can end up becoming the norm in communities if left unchecked.

Fighting back

Despite these sobering results, there are several ways this research can help us create better online spaces for public discussion.

By understanding what leads to trolling, we can now better predict when trolling is likely to happen.

This can let us identify potentially contentious discussions ahead of time and preemptively alert moderators, who can then intervene in these aggressive situations.

Machine learning algorithms can also sort through millions of posts much more quickly than any human. By training computers to spot trolling behavior, we can identify and filter undesirable content with much greater speed.

Social interventions can also reduce trolling. If we allow people to retract recently posted comments, then we may be able to minimize regret from posting in the heat of the moment. Altering the context of a discussion, by prioritizing constructive comments, can increase the perception of civility. Even just pinning a post about a community's rules to the top of discussion pages helps, as a recent experiment conducted on Reddit showed.

Nonetheless, there's lots more work to be done to address trolling. Understanding the role of organized trolling can limit some types of undesirable behavior.

Trolling also can differ in severity, from swearing to targeted bullying, which necessitates different responses.

It's also important to differentiate the impact of a troll comment from the author's intent: Did the troll mean to hurt others, or was he or she just trying to express a different viewpoint? This can help separate undesirable individuals from those who just need help communicating their ideas.

When online discussions break down, it's not just sociopaths who are to blame. We are also at fault. Many "trolls" are just people like ourselves who are having a bad day. Understanding that we're responsible for both the inspiring and depressing conversations we have online is key to having more productive online discussions.

Psychology **Facebook** **Social media** **Internet** **Media** **Reddit** **Trolling** **Comments** **CNN**