

# Five points for anger, one for a ‘like’: How Facebook’s formula fostered rage and misinformation

Facebook engineers gave extra value to emoji reactions, including ‘angry,’ pushing more emotional and provocative content into users’ news feeds

By [Jeremy B. Merrill](#) and [Will Oremus](#)

October 26, 2021 | Updated October 26, 2021 at 1:04 p.m. EDT



**Listen to article** 13 min



Five years ago, Facebook gave its users five new ways to react to a post in their news feed beyond the iconic “like” thumbs-up: “love,” “haha,” “wow,” “sad” and “angry.”

Behind the scenes, Facebook programmed the algorithm that decides what people see in their news feeds to use the reaction emoji as signals to push more emotional and provocative content — including content likely to make them angry. Starting in 2017, Facebook’s ranking algorithm treated emoji reactions as five times more valuable than “likes,” internal documents reveal. The theory was simple: Posts that prompted lots of reaction emoji tended to keep users more engaged, and keeping users engaged was the key to Facebook’s business.

Facebook’s own researchers were quick to suspect a critical flaw. Favoring “controversial” posts — including those that make users angry — could open “the door to more spam/abuse/clickbait inadvertently,” a staffer, whose name was redacted, wrote in one of the internal documents. A colleague responded, “It’s possible.”

The warning proved prescient. The company’s data scientists confirmed in 2019 that posts that sparked angry reaction emoji were disproportionately likely to include misinformation, toxicity and low-quality news.

That means Facebook for three years systematically amped up some of the worst of its platform, making it more prominent in users’ feeds and spreading it to a much wider audience. The power of the algorithmic promotion undermined the efforts of Facebook’s content moderators and integrity teams, who were fighting an uphill battle against toxic and harmful content.

The internal debate over the “angry” emoji and the findings about its effects shed light on the highly subjective human judgments that underlie Facebook’s news feed algorithm — the byzantine machine-learning software that decides for billions of people what kinds of posts they’ll see each time they open the app. The deliberations were revealed in disclosures made to the Securities and Exchange Commission and provided to Congress in redacted form by the legal counsel of whistleblower Frances Haugen. The redacted versions were reviewed by a consortium of news organizations, including The Washington Post.

“Anger and hate is the easiest way to grow on Facebook,” Haugen told the British Parliament on Monday.

In several cases, the documents show Facebook employees on its “integrity” teams raising flags about the human costs of specific elements of the ranking system — warnings that executives sometimes heeded and other times seemingly brushed aside. Employees evaluated and debated the importance of anger in society: Anger is a “core human emotion,” one staffer wrote, while another pointed out that anger-generating posts might be essential to protest movements against corrupt regimes.

An algorithm such as Facebook’s, which relies on sophisticated, opaque machine-learning techniques to generate its engagement predictions, “can sound mysterious and menacing,” said Noah Giansiracusa, a math professor at Bentley University in Massachusetts and author of the book “How Algorithms Create and Prevent Fake News.” “But at the end of the day, there’s one number that gets predicted — one output. And a human is deciding what that number is.”

Facebook spokesperson Dani Lever said: “We continue to work to understand what content creates negative experiences, so we can reduce its distribution. This includes content that has a disproportionate amount of angry reactions, for example.”

The weight of the angry reaction is just one of the many levers that Facebook engineers manipulate to shape the flow of information and conversation on the world’s largest social network — one that has been shown to influence everything from users’ emotions to political campaigns to atrocities.

Facebook takes into account numerous factors — some of which are weighted to count a lot, some of which count a little and some of which count as negative — that add up to a single score that the news feed algorithm generates for each post in each user’s feed, each time they refresh it. That score is in turn used to sort the posts, deciding which ones appear at the top and which appear so far down that you’ll probably never see them. That single all-encompassing scoring system is used to categorize and sort vast swaths of human interaction in nearly every country of the world and in more than 100 languages.

Facebook doesn’t publish the values its algorithm puts on different kinds of engagement, let alone the more than 10,000 “signals” that it has said its software can take into account in predicting each post’s likelihood of producing those forms of engagement. It often cites a fear of giving people with bad intentions a playbook to explain why it keeps the inner workings under wraps.

Facebook’s levers rely on signals most users wouldn’t notice, like how many long comments a post generates, or whether a video is live or recorded, or whether comments were made in plain text or with cartoon avatars, the documents show. It even accounts for the computing load that each post requires and the strength of the user’s Internet signal. Depending on the lever, the effects of even a tiny tweak can ripple across the network, shaping whether the news sources in your feed are reputable or sketchy, political or not, whether you saw more of your real friends or more posts from groups Facebook wanted you to join, or if what you saw would be likely to anger, bore or inspire you.

Beyond the debate over the angry emoji, the documents show Facebook employees wrestling with tough questions about the company's values, performing cleverly constructed analyses. When they found that the algorithm was exacerbating harms, they advocated for tweaks they thought might help. But those proposals were sometimes overruled.

When boosts, like those for emoji, collided with “deboosts” or “demotions” meant to limit potentially harmful content, all that complicated math added up to a problem in protecting users. The average post got a score of a few hundred, according to the documents. But in 2019, a Facebook data scientist discovered there was no limit to how high the ranking scores could go.

If Facebook's algorithms thought a post was bad, Facebook could cut its score in half, pushing most of instances of the post way down in users' feeds. But a few posts could get scores as high as a billion, according to the documents. Cutting an astronomical score in half to “demote” it would still leave it with a score high enough to appear at the top of the user's feed.

“Scary thought: civic demotions not working,” one Facebook employee noted.

The culture of experimentation ran deep at Facebook, as engineers pulled levers and measured the results. An experiment in 2012 that was published in 2014 sought to manipulate the emotional valence of posts shown in users' feeds to be more positive or more negative, and then observed whether their own posts changed to match those moods, raising ethical concerns, *The Post* reported at the time. Another, reported by Haugen to Congress this month, involved turning off safety measures for a subset of users as a comparison to see if the measures worked at all.

A previously unreported set of experiments involved boosting some people more frequently into the feeds of some of their randomly chosen friends — and then, once the experiment ended, examining whether the pair of friends continued communication, according to the documents. A researcher hypothesized that, in other words, Facebook could cause relationships to become closer.

In 2017, Facebook was trying to reverse a worrying decline in how much people were posting and talking to each other on the site, and the emoji reactions gave it five new levers to pull. Each emotional reaction was worth five likes at the time. The logic was that a reaction emoji signaled the post had made a greater emotional impression than a like; reacting with an emoji took an extra step beyond the single click or tap of the like button. But Facebook was coy with the public as to the importance it was placing on these reactions: The company told *Mashable* in 2017 that it was weighting them just “a little more than likes.”

The move was consistent with a pattern, highlighted in the documents, in which Facebook set the weights very high on new features it was trying to encourage users to adopt. By training the algorithm to optimize for those features, Facebook's engineers all but ensured they'd be widely used and seen. Not only that, but anyone posting on Facebook with the hope of reaching a wide audience — including publishers and political actors — would inevitably catch on that certain types of posts were working better than others.

At one point, CEO Mark Zuckerberg even encouraged users in a public reply to a user's comment to use the angry reaction to signal they disliked something, although that would make Facebook show similar content more often.

Replies to a post, which signaled a larger effort than the tap of a reaction button, were weighted even higher, up to 30 times as much as a like. Facebook had found that interaction from a user's friends on the site would create a sort of virtuous cycle that pushed users to post even more. The Wall Street Journal reported last month on how Facebook's greater emphasis on comments, replies to comments and replies to re-shares — part of a metric it called "meaningful social interactions" — further incentivized divisive political posts. (That article also mentioned the early weight placed on the angry emoji, though not the subsequent debates over its impact.)

The goal of that metric is to "improve people's experience by prioritizing posts that inspire interactions, particularly conversations, between family and friends," Lever said.

The first downgrade to the angry emoji weighting came in 2018, when Facebook cut it to four times the value of a like, keeping the same weight for all of the emotions.

But it was apparent that not all emotional reactions were the same. Anger was the least used of the six emoji reactions, at 429 million clicks per week, compared with 63 billion likes and 11 billion "love" reactions, according to a 2020 document. Facebook's data scientists found that angry reactions were "much more frequent" on problematic posts: "civic low quality news, civic misinfo, civic toxicity, health misinfo, and health antivax content," according to a document from 2019. Its research that year showed the angry reaction was "being weaponized" by political figures.

In April 2019, Facebook put in place a mechanism to "demote" content that was receiving disproportionately angry reactions, although the documents don't make clear how or where that was used, or what its effects were.

By July, a proposal began to circulate to cut the value of several emoji reactions down to that of a like, or even count them for nothing. The “angry” reaction, along with “wow” and “haha,” occurred more frequently on “toxic” content and misinformation. In another proposal, from late 2019, “love” and “sad” — apparently called “sorry” internally — would be worth four likes, because they were safer, according to the documents.

The proposal depended on Facebook higher-ups being “comfortable with the principle of different values for different reaction types,” the documents said. This would have been an easy fix, the Facebook employee said, with “fewer policy concerns” than a technically challenging attempt to identify toxic comments.

But at the last minute, the proposal to expand those measures worldwide was nixed.

“The voice of caution won out by not trying to distinguish different reaction types and hence different emotions,” a staffer later wrote.

Later that year, as part of a debate over how to adjust the algorithm to stop amplifying content that might subvert democratic norms, the proposal to value angry emoji reactions less was again floated. Another staffer proposed removing the button altogether. But again, the weightings remained in place.

Finally, last year, the flood of evidence broke through the dam. Additional research had found that users consistently didn’t like it when their posts received “angry” reactions, whether from friends or random people, according to the documents. Facebook cut the weight of all the reactions to one and a half times that of a like.

That September, Facebook finally stopped using the angry reaction as a signal of what its users wanted and cut its weight to zero, taking it out of the equation, the documents show. Its weight is still zero, Facebook’s Lever said. At the same time, it boosted “love” and “sad” to be worth two likes.

It was part of a broader fine-tuning of signals. For example, single-character comments would no longer count. Until that change was made, a comment just saying “yes” or “.” — tactics often used to game the system and appear higher in the news feed — had counted as 15 times the value of a like.

“Like any optimization, there’s going to be some ways that it gets exploited or taken advantage of,” Lars Backstrom, a vice president of engineering at Facebook, said in an emailed statement. “That’s why we have an integrity team that is trying to track those down and figure out how to mitigate them as efficiently as possible.”

But time and again, Facebook made adjustments to weightings after they had caused harm. Facebook wanted to encourage users to stream live video, which it favored over photo and text posts, so its weight could go as high as 600 times. That had helped cause “ultra-rapid virality for several low quality viral videos,” a document said. Live videos on Facebook played a big role in political events, including both the racial justice protests last year after the killing of George Floyd and the riot at the U.S. Capitol on Jan. 6.

Immediately after the riot, Facebook frantically enacted its “Break the Glass” measures on safety efforts it had previously undone — including to cap the weight on live videos at only 60. Facebook didn’t respond to requests for comment about the weighting on live videos.

When Facebook finally set the weight on the angry reaction to zero, users began to get less misinformation, less “disturbing” content and less “graphic violence,” company data scientists found. As it turned out, after years of advocacy and pushback, there wasn’t a trade-off after all. According to one of the documents, users’ level of activity on Facebook was unaffected.

#### **CORRECTION**

An experiment that sought to manipulate the emotional valence of posts shown in users’ feeds to be more positive or more negative, and then observed whether their own posts changed to match those moods, took place in 2012, not 2014. It was published in 2014. This article has been corrected.

## **Read the series: Facebook under fire**

The **Facebook Papers** are a set of internal documents that were provided to Congress in redacted form by Frances Haugen’s legal counsel. The redacted versions were reviewed by a consortium of news organizations, including The Washington Post.

The trove of documents show how **Facebook CEO Mark Zuckerberg** has, at times, contradicted, downplayed or failed to disclose company findings on the impact of its products and platforms.

The documents also provided new details of the social media platform's role in fomenting the **storming of the U.S. Capitol**. An investigation by ProPublica and The Washington Post found that Facebook groups swelled with **at least 650,000 posts attacking the legitimacy of Joe Biden's victory** between Election Day and Jan. 6.

Facebook engineers gave **extra value to emoji reactions, including 'angry,'** pushing more emotional and provocative content into users' news feeds.

### Read more from The Post's investigation:

[Key takeaways from the Facebook Papers](#)

Show more 

---

## MORE ABOUT FACEBOOK

 HAND CURATED

What is the 'metaverse'? Facebook says it's the future of the Internet.

News • October 28, 2021

You've decided to quit Facebook. Here's how to migrate your online life elsewhere.

News • November 10, 2021

Facebook privacy settings to change now

News • September 23, 2021

View 3 more stories 

---