

# “Know the Facts”: Do Social Media COVID-19 Banners Help?



Franzi Roesner

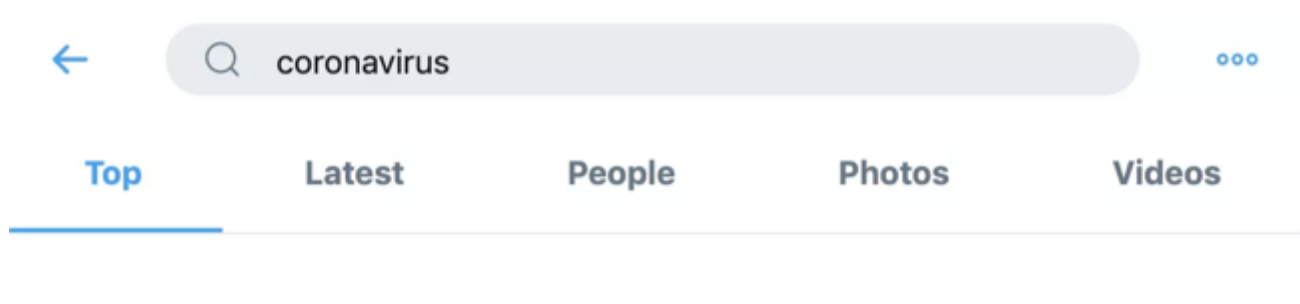
[Follow](#)

Apr 2, 2020 · 6 min read

*Authors: [Christine Geeng](#), [Jevin D. West](#), [Franziska Roesner](#)*

As everyone is desperate to make sense of how to handle and what will happen with COVID-19, significant amounts of misinformation are also spreading to fill the information vacuum. False (but believable) stories about topics such as [martial law in some US states](#), the [secret origins of the virus](#), and [dolphins appearing in the canals of Venice](#) abound. While some of these stories may be harmless (it’s nice to think about the dolphins!), others can cause significant harm as people make [dangerous decisions in attempts to protect themselves from the virus](#) or find themselves [targeted by racist attacks](#).

In the face of these swirling rumors, social media platforms have taken [nearly unprecedented strong stances against misinformation](#) posted on their platforms — much more concerted efforts than in the past. For example, Facebook and Twitter (and others) now show coronavirus information banners linking people to authoritative sources when they search. In attempts to combat misinformation more generally, Facebook also explicitly labels false fact-checked posts and Twitter labels manipulated media (e.g., videos).



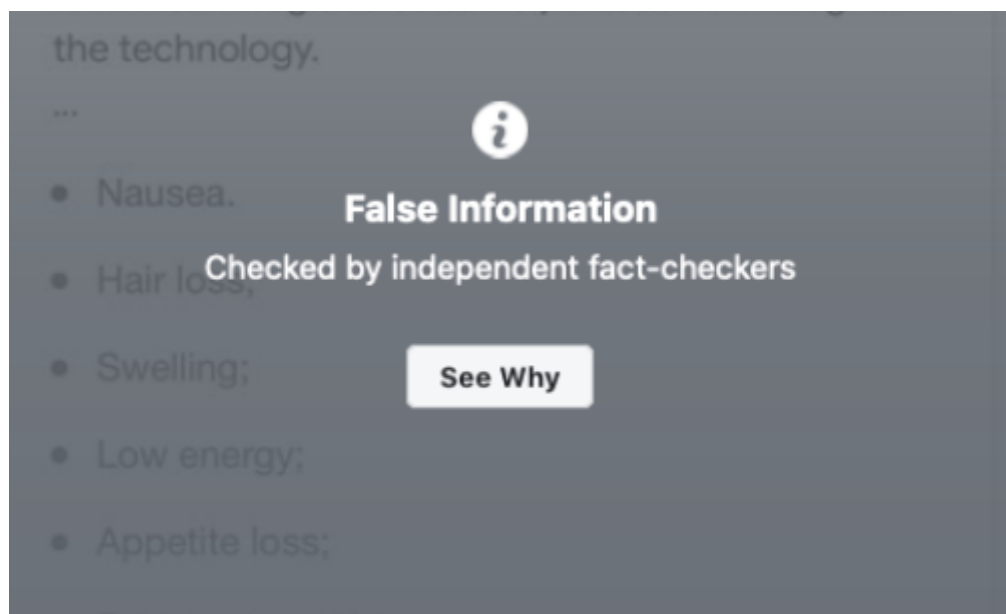
## Know the facts

To make sure you get the best information on the novel coronavirus, resources are available from the Centers for Disease Control and Prevention (CDC).

Centers for Disease Control and Prevention

@CDCGov

Twitter shows an authoritative banner, linking to the CDC, for U.S.-based searches about coronavirus.



Facebook grays out and labels posts as “false information” based on input from third-party fact checkers.

We are researchers at the [University of Washington](#), from the [Paul G. Allen School of Computer Science & Engineering](#) and the [Information School](#), and affiliated with the UW’s [Center for an Informed Public](#). In our [recent past work](#), we studied how people interact with misinformation on their social media feeds.

Given the current (mis)information environment, we now wanted to know: **What are people’s experiences with COVID-19 related misinformation *right now*? Are the social media platforms’ mitigation strategies helping?**

To answer these questions, we ran a survey from March 20–26, 2020, with approval from the University of Washington’s Human Subjects Review Board (IRB). We collected

203 responses via the crowdsourced survey platform Prolific and 110 additional responses by advertising the survey via our own social media accounts. We combined the two sets of survey results for this initial analysis.

The results from our survey suggest some preliminary conclusions. **We stress that this research has not yet been peer-reviewed**, but given the circumstances, we hope these early results are a useful addition to the conversation around the tech industry’s response to COVID-19 misinformation.

**1. Misinformation about COVID-19 is rampant.** 79.5% of participants reported having seen COVID-19 related misinformation on social media or elsewhere, and 33.9% reported believing something false themselves. The most commonly encountered falsehoods included that scientists have confirmed that the novel coronavirus came from eating bats ([fact check](#)), that the virus is a leaked biological weapon from a lab in Wuhan ([fact check](#)), that a coronavirus vaccine already exists ([fact check](#)), and that you can self-check whether you have COVID-19 by holding your breath for 10 or more seconds ([fact check](#)).

**2. Social media platform designs are significantly outweighed by other strategies (like intentional web searches) for debunking misinformation.** When asked how they learned that something they saw was false (whether or not they initially believed it), participants told us most frequently that they had conducted a web search (39.6% of 240 who answered this question), sought out trusted sources (37.1%), saw a correction in a social media comment (19.2%), or heard a correction from someone directly (12.1%). Only 4.2% learned something was false because the social media platform had labeled it as such. The majority (71.7%) of respondents indicated they “knew it wasn’t true”, though we cannot verify whether respondents’ baseline knowledge was correct. (Note that participants could select multiple responses.)

**3. Post-specific social media interventions are viewed as more helpful, and seem to be more effective, than generic interventions.** We also asked participants to subjectively evaluate how helpful social media interventions are, on a 5-point Likert scale from “Not at all helpful” (1) to “Extremely helpful” (5). We find that participants

tended to rate post-specific interventions as more helpful. For example, comparing the 30 participants who had seen both Facebook interventions, these participants found the post-specific "False Information" label significantly more helpful than the generic banner. The median helpfulness rating of the generic banner was 2 ("Slightly helpful"), and the median rating of the specific label was 4 ("Very helpful") (Wilcoxon signed-rank test,  $V = 4$ ,  $Z = -4.13$ ,  $p < 0.05$ ,  $r = 0.75$ ).

Considering effectiveness in addition to perception, only 13.3% of 105 participants who saw the Facebook banner said that they had ever clicked on it. Meanwhile, 32.3% of the 65 participants who saw the Facebook "False Information" label said that they no longer believe the content of the post due to the label. (We note that the number of participants who indicated that the label was effective when we asked explicitly about it was larger (21) than when we asked more generally about how they learned something was false (10). Participants may have been thinking of different experiences when answering the more general question, or forgotten about the label except when reminded of it explicitly.) 50.8% said that they had *already* not believed the false-labeled post, and only 6.2% said that they continued to believe the post, or believed it more, given the label.

The median helpfulness rating of the generic Twitter banner was 3 ("Somewhat helpful"), with a reported clickthrough rate of 32.8% by 58 participants who saw the banner. The difference in helpfulness rating between the Facebook and Twitter banners, for people who saw both, was not statistically significant. The post-specific "Manipulated media" label on Twitter also appears to perform better than the generic banner, but the number of participants who reported having seen this label is small.

**4. People have a range of subjective reactions to social media platform misinformation interventions.** While we are still analyzing the qualitative free-response data from our survey, our preliminary results suggest that participants' opinions about social media interventions range from the positive ("I thought it was good that Facebook was trying to do something to inform people better") to the neutral ("I didn't think much of it. I follow the news so I didn't click on this one because I already know the basic details") to the negative ("I don't like it. I don't need Facebook to tell me this, and I don't trust their automated way of detecting it") to (rarely) the hostile ("I was irritated because it is another in a long list of 'tools' to 'protect' users. In my opinion, this

label assumes people are morons and unable to discern what’s true, false and/or misleading”).

Overall, our initial results suggest that social media platform interventions are not yet doing the heaviest lifting when it comes to helping people avoid COVID-19 related (or other) misinformation — but that many people are receptive to these attempts.

Stepping back, we still have much more to learn about the role of social media platforms in combating misinformation and shaping behavior. For example, while the generic banners may not change behaviors in the moment, perhaps they are having a more subtle, sustained impact on how people evaluate information in their feeds? Moreover, since we ran our initial survey, some of these platforms have begun taking even stronger stances directly encouraging (or preventing discouragement of) social distancing behaviors (e.g., Facebook, Twitter). Are these interventions changing behaviors, or only reinforcing them among people who are already receptive? And what does the future hold — are certain platform-based interventions considered more appropriate during a situation like a global pandemic, and will changed norms persist afterwards? We look forward to conducting and learning about further research on these questions.

(With [Christine Geeng](#) and [Jevin West](#).)

[Covid 19](#) [Coronavirus](#) [Misinformation](#) [Facebook](#) [Twitter](#)

[About](#) [Help](#) [Legal](#)

Get the Medium app

