The catch to putting warning labels on fake news

March 3 2020, by Peter Dizikes

Credit: CC0 Public Domain

After the 2016 U.S. presidential election, Facebook began putting warning tags on news stories fact-checkers judged to be false. But there's a catch: Tagging some stories as false makes readers more willing to

believe other stories and share them with friends, even if those additional, untagged stories also turn out to be false.

That is the main finding of a new study co-authored by an MIT professor, based on multiple experiments with news consumers. The researchers call this unintended consequence—in which the selective labeling of false news makes other news stories seem more legitimate—the "implied-truth effect" in news consumption.

"Putting a warning on some content is going to make you think, to some extent, that all of the other content without the warning might have been checked and verified," says David Rand, the Erwin H. Schell Professor at the MIT Sloan School of Management and co-author of a newly published paper detailing the study.

"There's no way the fact-checkers can keep up with the stream of misinformation, so even if the warnings do really reduce belief in the tagged stories, you still have a problem, because of the implied truth effect," Rand adds.

Moreover, Rand observes, the implied truth effect "is actually perfectly rational" on the part of readers, since there is ambiguity about whether untagged stories were verified or just not yet checked. "That makes these warnings potentially problematic," he says. "Because people will reasonably make this inference."

Even so, the findings also suggest a solution: Placing "Verified" tags on stories found to be true eliminates the problem.

The paper, "The Implied Truth Effect," has just appeared in online form in the journal *Management Science*. In addition to Rand, the authors are Gordon Pennycook, an assistant professor of psychology at the University of Regina; Adam Bear, a postdoc in the Cushman Lab at

Harvard University; and Evan T. Collins, an undergraduate researcher on the project from Yale University.

BREAKING: More labels are better

To conduct the study, the researchers conducted a pair of online experiments with a total of 6,739 U.S. residents, recruited via Amazon's Mechanical Turk platform. Participants were given a variety of true and false news headlines in a Facebook-style format. The false stories were chosen from the website Snopes.com and included headlines such as "BREAKING NEWS: Hillary Clinton Filed for Divorce in New York Courts" and "Republican Senator Unveils Plan To Send All Of America's Teachers Through A Marine Bootcamp."

The participants viewed an equal mix of true stories and false stories, and were asked whether they would consider sharing each story on social media. Some participants were assigned to a control group in which no stories were labeled; others saw a set of stories where some of the false ones displayed a "FALSE" label; and some participants saw a set of stories with warning labels on some false stories and "TRUE" verification labels for some true stories.

In the first place, stamping warnings on false stories does make people less likely to consider sharing them. For instance, with no labels being used at all, participants considered sharing 29.8 percent of false stories in the sample. That figure dropped to 16.1 percent of false stories that had a warning label attached.

However, the researchers also saw the implied truth effect take effect. Readers were willing to share 36.2 percent of the remaining false stories that did not have warning labels, up from 29.8 percent.

"We robustly observe this implied-truth effect, where if false content

doesn't have a warning, people believe it more and say they would be more likely to share it," Rand notes.

But when the warning labels on some false stories were complemented with verification labels on some of the true stories, participants were less likely to consider sharing false stories, across the board. In those circumstances, they shared only 13.7 percent of the headlines labeled as false, and just 26.9 percent of the nonlabeled false stories.

"If, in addition to putting warnings on things fact-checkers find to be false, you also put verification panels on things fact-checkers find to be true, then that solves the problem, because there's no longer any ambiguity," Rand says. "If you see a story without a label, you know it simply hasn't been checked."

Policy implications

The findings come with one additional twist that Rand emphasizes, namely, that participants in the survey did not seem to reject warnings on the basis of ideology. They were still likely to change their perceptions of stories with warning or verifications labels, even if discredited news items were "concordant" with their stated political views.

"These results are not consistent with the idea that our reasoning powers are hijacked by our partisanship," Rand says.

Rand notes that, while continued research on the subject is important, the current study suggests a straightforward way that social media platforms can take action to further improve their systems of labeling online news content.

"I think this has clear policy implications when platforms are thinking about attaching warnings," he says. "They should be very careful to

check not just the effect of the warnings on the content with the tag, but also check the effects on all the other content."

More information: Gordon Pennycook et al, The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings, *Management Science* (2020). DOI: 10.1287/mnsc.2019.3478

Provided by Massachusetts Institute of Technology

Citation: The catch to putting warning labels on fake news (2020, March 3) retrieved 28 April 2024 from https://phys.org/news/2020-03-fake-news.html