

Warning labels from fact checkers work—even if you don't trust them—says study

September 3 2024



Experimental stimuli. Credit: *Nature Human Behaviour* (2024). DOI: 10.1038/s41562-024-01973-x

Do you trust fact-checkers? It might not matter. A new *Nature Human Behaviour* [paper](#) from MIT Sloan School of Management Ph.D. candidate Cameron Martel and professor David Rand reveals a surprising truth: fact-checker warning labels on social media can significantly reduce belief in and spread of misinformation, even among those who harbor doubts about the fact-checkers themselves.

Rumors and falsehoods can spread quickly on social media, making it difficult for users to separate fact from fiction. In response, most major platforms have partnerships with third-party fact-checking organizations and attach warning labels to content found to be false or misleading, an approach that Martel and Rand's previous research suggests works on average.

However, [trust](#) in fact-checkers is not universal or consistent across the [political spectrum](#)—and neither is exposure to misinformation. In the United States, [research](#) has shown that political conservatives are more likely to see and share misinformation and [less likely to trust fact-checking](#), raising concerns that these interventions could potentially backfire.

"Most people don't see much misinformation," explained Martel. "And if the people who are more likely to be exposed to misinformation are less likely to trust fact-checkers, it's important to understand whether warning labels are effective for that group."

Measuring mistrust in fact-checkers

To answer those questions, Martel and Rand used a two-part approach. First, they conducted a correlational study to validate a measure of trust in fact-checkers and identify correlates of mistrust.

In line with previous studies, the researchers found that [Republican-leaning survey participants were less likely to trust fact-checkers](#)—regardless of whether the fact-checking organizations skewed right or left. They also saw that other traits interacted significantly with respondents' political affiliations to shape their attitudes.

Republican respondents who knew more about news production, who scored more highly on a cognitive reasoning test, and who had higher

web use skills were even less trusting of fact-checkers. These factors did not predict differences among Democratic respondents. However, higher self-reported digital media literacy was correlated with increased trust in fact-checkers, regardless of political affiliation.

Attitudes vs. actual responses

Next, Martel and Rand conducted a series of experiments with over 14,000 participants across the United States to test how media warning labels impacted responses to false headlines. Participants were exposed to a mix of politically balanced true and false headlines. Participants either saw most false headlines accompanied by warning labels similar to those used by Facebook, or no warning labels at all. Participants then rated the accuracy of each [headline](#) or indicated their willingness to share it.

While warning labels were somewhat more effective for individuals who scored higher on trust in fact-checkers, they also consistently and significantly reduced belief in and willingness to share false headlines among participants who demonstrated distrust in fact-checkers. This held true even among participants who scored in the bottom quartile for trust in fact-checkers.

"Misinformation warning labels worked for even the respondents in our sample who were least trusting of fact-checkers and farthest right politically—and we saw no evidence of any backfire effect," said Rand. "This research builds on our existing body of work demonstrating the efficacy of warning labels, and gives us reassurance that their impacts aren't one-sided."

So, what explains the discrepancy between users' attitudes toward fact-checkers and their response to [warning](#) labels?

Martel and Rand point to several possible explanations. The labels could prompt a more critical evaluation of the headlines, or individuals might refrain from sharing or expressing belief in headlines labeled as false because of the risk of reputational harm. It's also possible that politically engaged Republicans are more likely to express distrust in fact-checkers because they have been cued to do so, rather than because of a deeply held belief.

Regardless of the reason, says Martel, the research findings are good news for those concerned about the [spread of misinformation](#).

"Labels aren't perfect," he noted. "It's important that platforms also have other options, like downranking or removal, for content that is potentially more harmful. However, this work shows that content warnings are a useful tool that can work for a broad range of people, even if they say they don't trust them."

More information: Cameron Martel et al, Fact-checker warning labels are effective even for those who distrust fact-checkers, *Nature Human Behaviour* (2024). [DOI: 10.1038/s41562-024-01973-x](https://doi.org/10.1038/s41562-024-01973-x)

Provided by MIT Sloan School of Management

Citation: Warning labels from fact checkers work—even if you don't trust them—says study (2024, September 3) retrieved 5 September 2024 from <https://phys.org/news/2024-09-fact-checkers-dont.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.