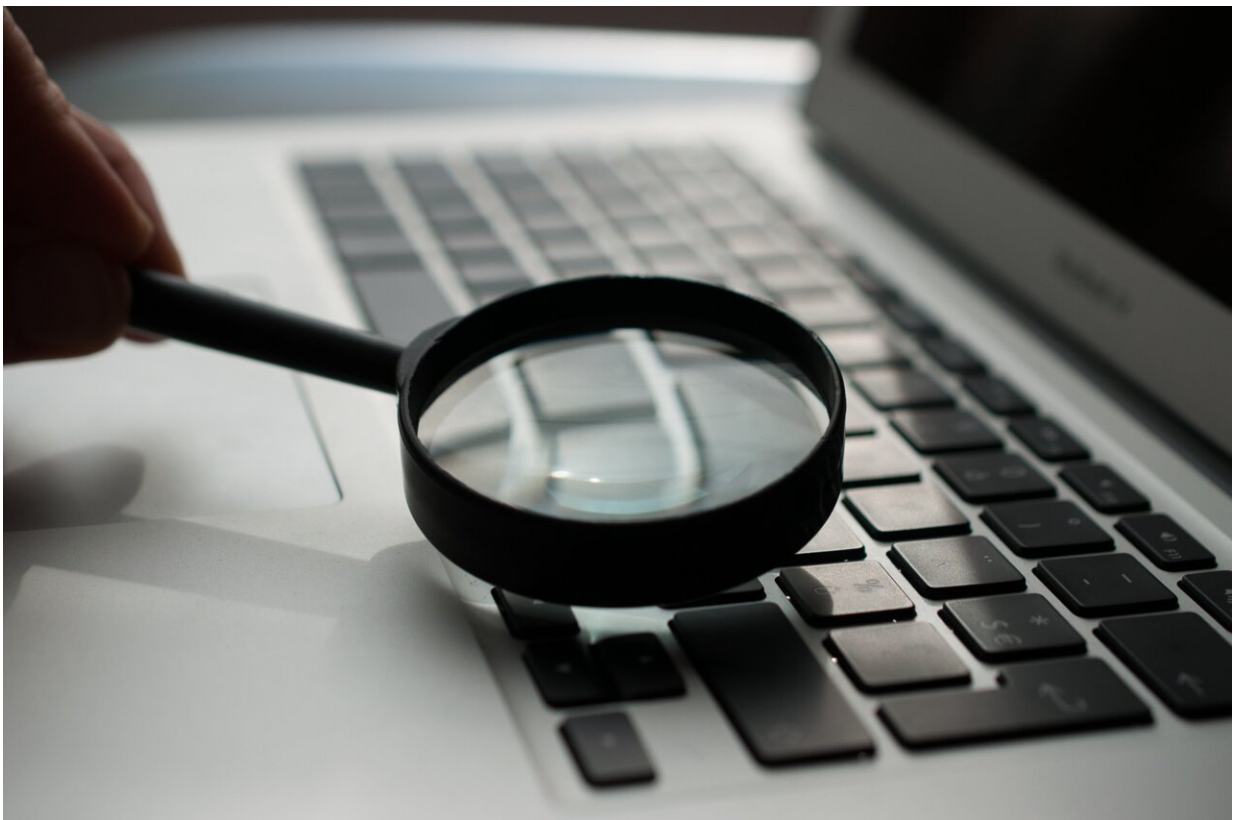# Experiment shows groups of laypeople reliably rate stories as effectively as fact-checkers do

September 1 2021, by Peter Dizikes



Credit: Unsplash/CC0 Public Domain

In the face of grave concerns about misinformation, social media networks and news organizations often employ fact-checkers to sort the

real from the false. But fact-checkers can only assess a small portion of the stories floating around online.

A new study by MIT researchers suggests an alternate approach: Crowdsourced accuracy judgements from groups of normal readers can be virtually as effective as the work of professional fact-checkers.

"One problem with fact-checking is that there is just way too much content for professional fact-checkers to be able to cover, especially within a reasonable time frame," says Jennifer Allen, a Ph.D. student at the MIT Sloan School of Management and co-author of a newly published paper detailing the study.

But the current study, examining over 200 news stories that Facebook's algorithms had flagged for further scrutiny, may have found a way to address that problem, by using relatively small, politically balanced groups of lay readers to evaluate the headlines and lead sentences of news stories.

"We found it to be encouraging," says Allen. "The average rating of a crowd of 10 to 15 people correlated as well with the fact-checkers' judgments as the fact-checkers correlated with each other. This helps with the scalability problem because these raters were regular people without fact-checking training, and they just read the headlines and lead sentences without spending the time to do any research."

That means the crowdsourcing method could be deployed widely—and cheaply. The study estimates that the cost of having readers evaluate news this way is about $0.90 per story.

"There's no one thing that solves the problem of false news online," says David Rand, a professor at MIT Sloan and senior co-author of the study. "But we're working to add promising approaches to the anti-

misinformation tool kit."

The paper, "Scaling up Fact-Checking Using the Wisdom of Crowds," is being published today in *Science Advances*. The co-authors are Allen; Antonio A. Arechar, a research scientist at the MIT Human Cooperation Lab; Gordon Pennycook, an assistant professor of behavioral science at University of Regina's Hill/Levene Schools of Business; and Rand, who is the Erwin H. Schell Professor and a professor of management science and brain and cognitive sciences at MIT, and director of MIT's Applied Cooperation Lab.

## A critical mass of readers

To conduct the study, the researchers used 207 news articles that an internal Facebook algorithm identified as being in need of fact-checking, either because there was reason to believe they were problematic or simply because they were being widely shared or were about important topics like health. The experiment deployed 1,128 U.S. residents using Amazon's Mechanical Turk platform.

Those participants were given the headline and lead sentence of 20 [news stories](#) and were asked seven questions—how much the story was "accurate," "true," "reliable," "trustworthy," "objective," "unbiased," and "describ[ing] an event that actually happened"—to generate an overall accuracy score about each news item.

At the same time, three professional fact-checkers were given all 207 stories —asked to evaluate the stories after researching them. In line with other studies on fact-checking, although the ratings of the fact-checkers were highly correlated with each other, their agreement was far from perfect. In about 49 percent of cases, all three fact-checkers agreed on the proper verdict about a story's facticity; around 42 percent of the time, two of the three fact-checkers agreed; and about 9 percent of the

time, the three fact-checkers each had different ratings.

Intriguingly, when the regular readers recruited for the study were sorted into groups with the same number of Democrats and Republicans, their average ratings were highly correlated with the professional fact-checkers' ratings—and with at least a double-digit number of readers involved, the crowd's ratings correlated as strongly with the fact-checkers as the fact-checkers' did with each other.

"These readers weren't trained in fact-checking, and they were only reading the headlines and lead sentences, and even so they were able to match the performance of the fact-checkers," Allen says.

While it might seem initially surprising that a crowd of 12 to 20 readers could match the performance of professional fact-checkers, this is another example of a classic phenomenon: the wisdom of crowds. Across a wide range of applications, groups of laypeople have been found to match or exceed the performance of expert judgments. The current study shows this can occur even in the highly polarizing context of misinformation identification.

The experiment's participants also took a political knowledge test and a test of their tendency to think analytically. Overall, the ratings of people who were better informed about civic issues and engaged in more analytical thinking were more closely aligned with the fact-checkers.

"People that engaged in more reasoning and were more knowledgeable agreed more with the fact-checkers," Rand says. "And that was true regardless of whether they were Democrats or Republicans."

## Participation mechanisms

The scholars say the finding could be applied in many ways—and note

that some social media behemoths are actively trying to make crowdsourcing work. Facebook has a program, called Community Review, where laypeople are hired to assess news content; Twitter has its own project, Birdwatch, soliciting reader input about the veracity of tweets. The wisdom of crowds can be used either to help apply public-facing labels to content, or to inform ranking algorithms and what content people are shown in the first place.

To be sure, the authors note, any organization using crowdsourcing needs to find a good mechanism for participation by readers. If participation is open to everyone, it is possible the crowdsourcing process could be unfairly influenced by partisans.

"We haven't yet tested this in an environment where anyone can opt in," Allen notes. "Platforms shouldn't necessarily expect that other crowdsourcing strategies would produce equally positive results."

On the other hand, Rand says, news and social media organizations would have to find ways to get a large enough groups of people actively evaluating news items, in order to make the crowdsourcing work.

"Most people don't care about politics and care enough to try to influence things," Rand says. "But the concern is that if you let people rate any content they want, then the only people doing it will be the ones who want to game the system. Still, to me, a bigger concern than being swamped by zealots is the problem that no one would do it. It is a classic public goods problem: Society at large benefits from people identifying misinformation, but why should users bother to invest the time and effort to give ratings?"