

# Finding a vaccine for misinformation

*With deliberate deception a growing threat online, social scientists are devising ways to fight back with “cognitive inoculations”*

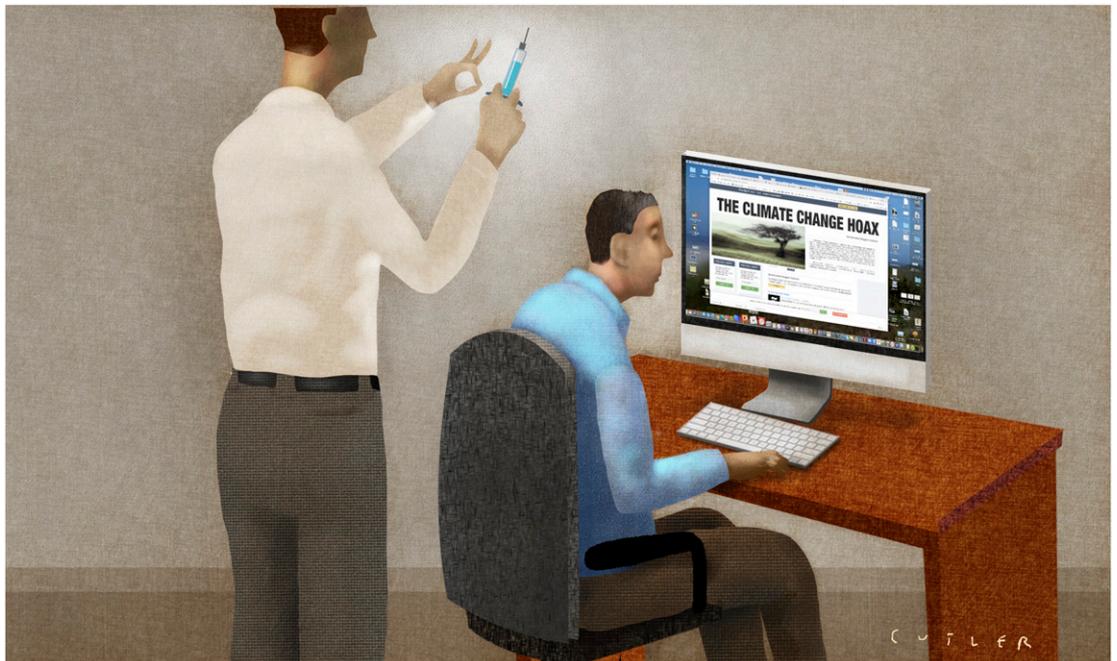
**Gayathri Vaidyanathan**, *Science Writer*

In early March, after a wave of coronavirus cases struck a Muslim congregation in India, the hashtag #CoronaJihad went viral on Indian Twitter, and Islamophobic messages began to surge on social media. In one case, a false video showed purportedly Muslim men licking plates—allegedly to spread the novel coronavirus (COVID-19). In reality, these men belonged to a community that strongly discourages wasting food. No one knows who crafted these false messages, or why. But anti-Muslim attacks increased after their release.

It’s hardly an isolated example. Misinformation has been rife during the pandemic—ranging from rumors about the virus being an escaped bioweapon to specious reports of a miracle cure doctors won’t tell you about. Of course, this is only the latest flavor of

fakery, which includes high-profile efforts such as climate-change denialism, antivaccine agitation (1), and Russian attempts to erode trust in the 2016 election. In every case, says Emma Spiro, a sociologist who is studying coronavirus misinformation at the University of Washington in Seattle, the risk is that “people may use the false information as the basis for decision making and actions that endanger themselves and others.”

Experts sometimes distinguish between *misinformation*, which is simply wrong and may even be an honest mistake, and *disinformation*, which is formulated with an intent to deceive. Whatever the name, stemming the onslaught can seem like a losing battle. Researchers, tech companies, journalists, and fact checkers have been



Hoping to fight an epidemic of misinformation, some researchers have demonstrated the promise of “inoculating” people with training videos and games even before those people are exposed to misinformation. Image credit: Dave Cutler (artist).

Published under the [PNAS license](#).

First published July 22, 2020.

trying to debunk misinformation for years. But researchers have found that this is not as simple as providing people with correct information and hoping it will supplant false beliefs. “What do you do if you have this mental model, and somebody tells you part of it is wrong?” notes Stephan Lewandowsky, a chair of cognitive psychology who’s seeking solutions at the University of Bristol in the United Kingdom. “It’s like ripping out some important part of what you think is happening in the world.”

Still, the fight against misinformation isn’t hopeless. One key may be finding a different line of attack. Some researchers have gotten promising, if preliminary, results by “inoculating” people with training videos and games even before they’re exposed to misinformation. Others are trying to deal with the aftermath of such exposure using approaches that change the minds of people who have already received a false message.

These studies are small so far and still in progress. But if the approaches work and they could be deployed on a large scale, the approach could slow chains of transmission of misinformation, researchers say. When elections are being decided with margins of a few percentage points, notes Sander van der Linden, a social psychologist at Cambridge University in the United Kingdom, even a small effect could make a huge difference.

### Psychology of Misinformation

Social scientists have identified at least three key factors governing how people absorb information, form beliefs, and modify behaviors—none of them having much to do with objective truth.

The first factor is social contagion, a formal name for people’s tendency to think and act like their friends and family. Researchers have long documented the effect of these linkages on smoking, obesity, and a variety of other behaviors. Initially, they thought behaviors spread through social networks simply by jumping from person to person the way a real virus does. But some practices, such as the use of a mosquito net in malaria-prone regions, don’t catch on easily, leaving researchers puzzled.

In 2010, in an effort to understand what’s really going on, Damon Centola, director of the network dynamics group at the University of Pennsylvania in Philadelphia, set up an experiment with 1,528 test subjects split between two social networks. In one network, subjects were clustered into “neighborhoods” with loose interconnections between them. In the other network, the subjects were randomly associated across the network, much like they would be in the case of viral disease spread. Centola also assigned each participant a few “health buddies,” meant to mimic family and friends (2). Then he released information about a new health forum to both the networks and watched as people began joining the forum. Each time a person joined, their health buddies were notified.

Centola found that people were more likely to join the forum if many of their buddies had already joined. He likewise found that the behavior spread more quickly in the neighborhood network than the random-

association network. Taken together, these findings suggested to Centola that a person adopts a behavior only after reinforcement from multiple people, which happens sooner in the neighborhood network, and only after judging the people in one’s social network who do and do not advocate for a belief or behavior. “It’s an incredibly sophisticated social experience,” says Centola, who calls this model of behavior spread a “complex contagion.”

Another key factor is how messages are framed to evoke deeper narratives that already exist in the listeners’ minds. These cause-and-effect storylines can be especially powerful when they evoke negative emotions such as fear, anger, or disgust—even when the story is false. Take for example the following specious statement: “A medical expert working for the government found a causal link between vaccines and autism, but federal lawmakers influenced by the powerful pharmaceutical lobby helped bury the info.” Such messages are said to be “structurally coherent”: they are easy to grasp and recall. A coherent story works because our minds don’t just encode facts and events into memory, says Valerie Reyna, a psychologist at Cornell University in Ithaca, NY, and director of the university’s Human Neuroscience Institute. We also store the bottom-line meaning, or “gist”—and it is the stored gist, not the facts, that typically guides our beliefs and behaviors.

To demonstrate the importance of coherence, Reyna and David Broniatowski, director of the Decision Making and Systems Architecture Laboratory at George Washington University in Washington, DC, collected 10,000 tweets about vaccines from 2014 to 2017 (3). Then they looked for commonalities among the 46% that were retweeted. Quite consistently, they found that the retweeted posts had a strong (but false) bottom-line message about vaccines causing autism. Tweets that contained facts and statistics did not spread far.

But which gists resonate with us and which do we encode? Some scientists say that depends on the third factor: our “worldview”—what Reyna describes as preexisting internal stories based on our mental tapestry of culture, knowledge, beliefs, and life experiences. For example, a news article that states, “Flu shot induces 4.4-fold increase in nonflu acute respiratory infections” might lead a nonbeliever to store the gist “don’t get the flu shot,” whereas an expert might store something like, “there are problems with reporting bias in the data collection, and the report is not trustworthy.”

These internal stories, or narratives, are why straightforward fact checks and debunking of dubious information may not help. “Any direct sort of attempt to say, ‘Hey, you’re wrong, you need to think differently’ [doesn’t work] because people don’t like that,” says Lewandowsky. A correction that contradicts a deeply held narrative can easily backfire and make people hold on to the original information more strongly than ever (4). And this resistance to fact checks can become especially strong when ideology is involved. According to a controversial theory of

“cultural cognition” formulated by sociologist Dan Kahan of Yale University in New Haven, CT, those Americans who prize individualism, free enterprise, and deregulation tend to dismiss the risks of climate change even when they fully understand that the science says otherwise (5). People think they have to hold strongly to their views or risk losing their status in their social circle.

### Tackling Misinformation

Taking into account social contagion, framing, and worldview—and recognizing that researchers are still divided over how much each factor matters—what strategies, then, can best protect people against misinformation?

One approach dates back to the 1960s, when Columbia University (New York) social psychologist William McGuire found that people could be “inoculated” against attacks on cultural truisms such as “tooth brushing after every meal is a good idea” (6).

**“Nobody likes to be duped by fake news, and part of the inoculation is warning people that there are people trying to mislead them.”**

**—Sander van der Linden**

Just as a conventional vaccine uses a weakened or dead form of a virus to prime the immune system against future infections, McGuire’s intervention helped people develop a sort of resistance by first delivering a weak attack—“too much tooth brushing wears down the enamel,” for instance—then following it with a direct refutation: “soft toothbrushes clean the teeth without producing any wear.”

Lewandowsky and van der Linden are trying to deploy a similar kind of “cognitive inoculation” in much more controversial arenas, such as climate denial. In work published in 2017, Lewandowsky and colleagues exposed people to a message that detailed misleading arguments used by the tobacco industry to sow confusion about the health effects of tobacco (7)—corporate messaging tactics that resemble those of climate deniers. The researchers explained why these arguments are flawed, and then they gave the subjects a false news story stating that some researchers doubt humans cause climate change. They found that participants exposed to this sort of “cognitive vaccine” were more skeptical of the false claim.

Since then, Lewandowsky has been trying to develop truth-adopting interventions that can work on all kinds of misinformation. His idea is to show in general how arguments that don’t fit together are incoherent and must be scientifically wrong. “The earth isn’t both flat and round, it’s one or the other,” he says.

In an upcoming study, Lewandowsky describes a set of 2- to 3-minute training videos that he’s developed to help people detect faulty arguments, and he shows that the videos help subjects spot such arguments in the

messages of climate deniers and antivaxxers. He hopes to begin deploying his videos using Facebook ads and then move to a large-scale platform such as YouTube. “In society at large, if 1 in 100 people is less likely to share fake news, that’ll make a difference because it cuts down on the spread of that stuff,” he says.

### Bad News

Meanwhile, van der Linden and his colleagues have been developing a similar inoculation protocol using a different vaccine message (8). Their results suggest that most Democrats, Independents, and Republicans all respond to inoculation and are protected to a similar degree. This outcome challenges a theory of cultural cognition that predicts that Republicans would be ideologically motivated to reject arguments that don’t fit their worldviews. Van der Linden contends that most people are susceptible to misinformation not because they hold strong worldviews but rather because they do not think carefully about the underlying logic.

This belief has led him to make his own attempt a broad-spectrum approach that could hopefully work against all kinds of misinformation. “Nobody likes to be duped by fake news, and part of the inoculation is warning people that there are people trying to mislead them,” he says. But instead of just telling people about these malefactors, van der Linden and his team show them: They ask subjects to become purveyors of fake news in an online game called *Bad News*, which they released in 2018.

When the game starts, the players enter a virtual environment that includes an entire simulated social media ecosystem. There they are presented with sets of fake Tweets, images, and other content that they can use to rack up as many social media followers as possible, gain credibility, and sow chaos. At the first level, a player sets up a fake news website inside the game and earns an “impersonation” badge. At the second level, she chooses content that appeals to base emotions such as fear and anger, earning herself an “emotion” badge. At the third level, she learns to amplify Twitter scandals to drive a wedge between people in her simulated audience, which earns her a “polarization” badge. Next, she earns a “conspiracy” badge by crafting a lie that sows doubt on a credible institution such as the World Health Organization. Then, when her fake news outlet is attacked by a fact-checking site, she launches a counterassault to earn a “discredit” badge. Finally, she uses all that she has learned to generate a conspiracy that remakes history and earns her a “trolling” badge. At the end, the researchers administer a survey to again test the player’s susceptibility to misinformation.

In January, van der Linden’s team released results from a randomized, controlled trial showing that *Bad News* improved subjects’ ability to spot misinformation techniques and made them more confident in their judgments of fake news (9). The researchers have made the game available online; so far, about a million people have played it. Van der Linden says that the game confers at least partial “immunity” to many of the players.

## Changing Minds

As promising as these results are, though, neither van der Linden nor Lewandowsky think they are the whole answer. “There’s always going to be audiences that are relatively immune to inoculation,” says van der Linden.

These “inoculation-resistant” individuals are being addressed by Joshua Introne, a professor of information studies at Syracuse University, NY. He is recruiting people with pro- and antivaccine views, as determined by a preliminary survey, and asking them to play a web-based game he calls *The Story Loom*. The game starts by placing players randomly in a social network, giving them a few snippets of information drawn from a larger pool of antivaccine storylines, and asking them to organize the snippets into a story. Each person can also see and use the snippets of their immediate neighbors. The snippets can include information such as “the little girl is given a vaccine,” “the little girl gets sick,” or “the little girl has allergies.” The researchers can later provide snippets that retract or refute any of the information to see whether a correction reduces the influence of misinformation. In each round of the game, players build on their stories until they run out of time. Points are awarded for the number of snippets used and—to make sure people don’t just put the snippets in some random sequence—docked if the snippets are out of order.

By watching his subjects assemble story snippets that then spread across the Story Loom, Introne can monitor coherence—how well the story fits with a player’s background narrative—and learn how much coherence matters in making a story go viral. He gauges this fit, and coherence, using a mathematical model of information diffusion that uses the data collected from the game. This allows him to calculate the coherence of each snippet within a story, as well as how it jibes with the background antivaccine belief state of the player. And once corrections are deployed, he can gauge the impact of the correction on coherence and social spread.

For example, suppose that the players are presented with this piece of misinformation during the game: “A pharmaceutical company whistleblower says researchers falsified data to mislead people

about the efficacy of flu vaccines in the 1990s.” Then the players are presented with a correction like this: “The whistleblower is on trial for stealing industry secrets, and the Centers for Disease Control and Prevention (CDC) say there is no record of the company falsifying data.” But this appeal to authority could backfire, says Introne, because antivaxxers don’t typically believe the CDC to be trustworthy. A better correction would be to give an alternate account of the whistleblower’s behavior that doesn’t reinforce the antivaccine narrative—but that doesn’t conflict with it, either. For example: “Dr. Barnsworth, a 64-year-old lab technician at Merck, has been accused of intellectual property theft. He was caught attempting to sell digital records to undercover agents in an online chat forum. Barnsworth had been under surveillance because of a string of thefts in the lab where he worked.”

The game is still in development, but Introne believes that he can use this approach to target the weakest links in false narratives and bring people closer to changing their minds (10). He says that if he can deliver information that doesn’t conflict with a person’s belief state but still brings them around to a more accurate point of view, “then I’ve got a pretty powerful thing.”

With the COVID-19 pandemic still raging, there is an especially critical need, says Spiro, for effective ways to address misinformation. Whether it’s a bogus cure, such as claims that high doses of vitamin C, honey, and colloidal silver will cure coronavirus, or a rumor about a community spreading the disease, the messages follow the misinformation toolkit: They play on fear, have strong causal structures, or work well on people with particular worldviews. Already, van der Linden’s team is assembling a version of the *Bad News* game tailored to COVID-19 that can be deployed through WhatsApp.

But Spiro doesn’t foresee one universal cure-all. “There are strategies people can practice,” she says, “but it is also important to recognize that we will all make mistakes and share misinformation at some point.”

- 1 D. A. Broniatowski et al., Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am. J. Public Health* **108**, 1378–1384 (2018).
- 2 D. Centola, The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
- 3 D. Broniatowski, V. Reyna, To illuminate and motivate: A fuzzy-trace model of the spread of information online. *Comput. Math. Organ. Theory* 10.1007/s10588-019-09297-2 (2019). Accessed 14 May 2020.
- 4 M. S. Chan, C. R. Jones, K. Hall Jamieson, D. Albarracín, Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol. Sci.* **28**, 1531–1546 (2017).
- 5 D. Kahan, Fixing the communications failure. *Nature* **463**, 296–297 (2010).
- 6 W. McGuire, Resistance to persuasion conferred by active and passive refutation of the same and alternate counterarguments. *J. Abnorm. Soc. Psychol.* **63**, 326–332 (1961).
- 7 J. Cook, S. Lewandowsky, U. K. H. Ecker, Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS One* **12**, e0175799 (2017).
- 8 S. van der Linden, A. Leiserowitz, S. Rosenthal, E. Maibach, Inoculating the public against misinformation about climate change. *Glob. Chall.* **1**, 1600008 (2017).
- 9 M. Basol, J. Roozenbeek, S. van der Linden, Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *J. Cogn.* **3**, 2 (2020).
- 10 N. Walter, S. Murphy, How to unring the bell: A meta-analytic approach to correction of misinformation. *Commun. Monogr.* **85**, 423–441 (2018).