

How Science Gets Drawn Into Global Conspiracy Narratives

MARC TUTERS, TOM WILLAERT, AND TRISHA MEYER

A few short years ago, mRNA (messenger ribonucleic acid) was the subject of fundamental research, but it is now known as the basis for COVID-19 vaccines. At the same time, the concept has become linked—particularly on social media—to global conspiracy theories attributing nefarious motives to people associated with science. How did this happen?

In our work, we use social media data to track evolving narratives empirically. By analyzing the terms that have become associated with mRNA on Twitter since early 2020, we have gained insight into how seemingly innocuous scientific concepts acquire sinister connotations through association. Understanding how this process occurs can be helpful in determining which countermeasures might be effective.

Hashtags are key to this analysis. Used to cross-link social media posts, hashtags generally consist of a word preceded by a pound symbol: #mRNA, for example. Hashtags make such concepts easier to find because they can be easily searched. By observing how hashtags co-occur over time, we learn how ideas are linked to each other on social media. This approach is useful for understanding the ways in which disparate concepts become related to evolving narratives.

To find out how the term *mRNA* became connected to far-flung conspiracy theories, we collected a sample of 87,000 tweets containing the hashtag #mRNA over the three-year period from early 2020 to the end of 2022. This allowed us to look at how mRNA was juxtaposed with other ideas on social media over that time. Our analysis looks at time continuously, but we've found it helpful to take "slices" from the dataset to highlight the way the narrative took shape and then shifted over time.

We looked at where #mRNA occurred next to other hashtags, which gives a sense of how the term became connected to

other ideas. We presented this data visually, displaying the connections as a network where each node represents a different hashtag and each edge represents the number of times two hashtags co-occur in our dataset. Starting with tweets using the hashtag #mRNA, this method allows us to see how sometimes unexpected semantic networks of associations can develop around ideas. Although co-occurring hashtags should not be taken as representing general discussions about mRNA, their changing patterns over time may offer insights into how issues may be hijacked and misinformation spread.

In our first sample, from early 2020, the hashtags co-occurring with #mRNA were largely scientific or financial, reflecting prepandemic views. To make this network graph more readable, we "cleaned" the data, systematically removing all hashtag nodes above and below certain thresholds determined by number of connections. The distance between nodes indicates how often terms are used together in posts, and the size of the text reflects the sum of its connections. Thus, larger text shows terms that are highly interconnected. The color coding reflects which communities are involved, which is discovered through an automated process that examines connections. The first figure depicts a mostly pale green community made up of hashtags corresponding to scientific terms as well as a gray colored community devoted to discussing biotech investment.

These figures make it possible to examine how the narrative around mRNA evolved over time. As vaccines went into production, new semantic networks associated with the term quickly began to develop. One cannot simply assume this method represents all the opinions that are "out there" in society, but it can nevertheless be quite helpful in

understanding how dramatic conspiracy narratives attach themselves to aspects of science and technology.

In December 2020, around when Pfizer’s COVID-19 vaccine was given an emergency use authorization for widespread use, the term *mRNA* started to be associated with resistance to vaccines. By early 2021, the hashtag #mRNA started to be combined in Germany with other hashtags that reference concerns responding to “compulsory vaccination” (#impflicht), with charges of “crimes against humanity” and calls for a “Nuremberg 2.”

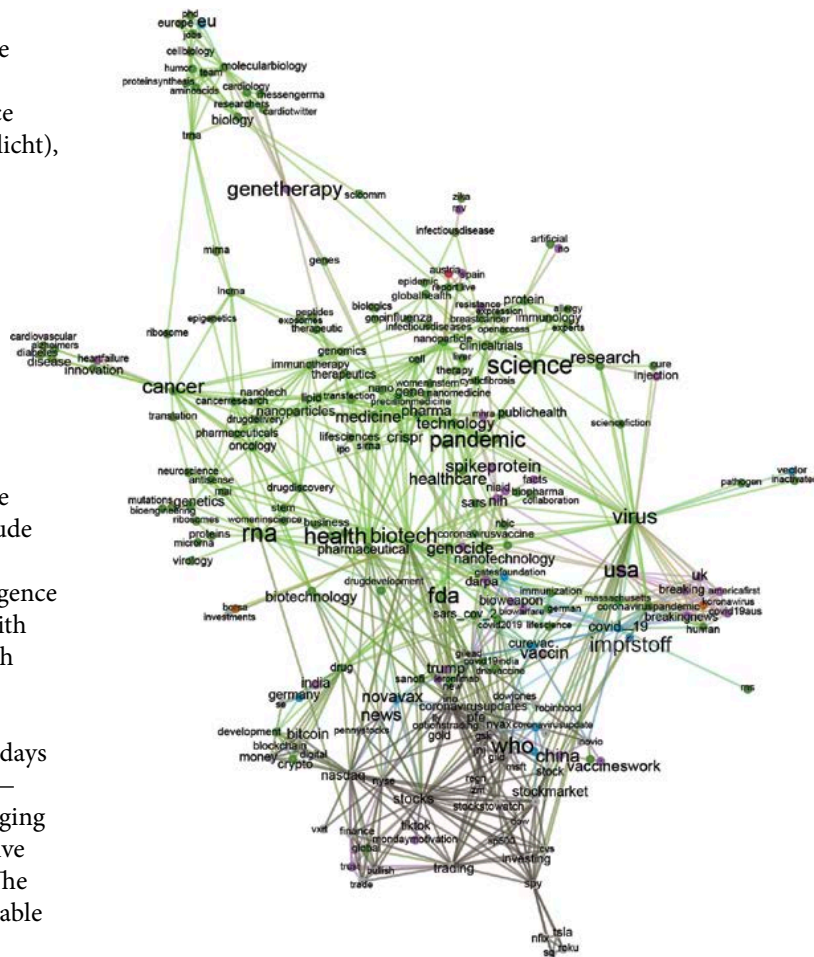
In Figure 2, you can see a blue community representing German hashtags that were not present the year before. These hashtags reflect language used in the anti-lockdown protests that were taking place in Germany at the time. The protests represented a wide range of ideologies, blending right-wing populism, left-wing libertarianism, and alternative health advocacy—the so-called lateral thinkers or *Querdenker* movement.

In addition to the large German community, you can see much smaller national grey and red communities that include hashtags connected to the anti-lockdown protests in the Netherlands and Italy, respectively. Also visible is the emergence of another large English-language community in purple, with hashtags in English referring to people and institutions such as Anthony Fauci and the Centers for Disease Control and Prevention, as well as #plandemic and #covid1984.

This visualization also provides a glimpse into the early days of the conspiracy that came to be called “The Great Reset”—through the hashtags #greatreset and #WEF. This wide-ranging and convoluted conspiracy concerns an eponymous initiative launched by the World Economic Forum (WEF) in 2020. The WEF’s Great Reset sought to promote a vision for a sustainable postpandemic world rooted in “a new social contract that honors the dignity of every human being.” Coming at a moment of rising populist sentiment, the WEF’s vague yet ambitious Great Reset plan fed directly into existing fears about craven “globalist” elites undermining both national sovereignty and “individual sovereignty,” a political stance that motivated many anti-vaccine activists. In late 2020, around the time of the US presidential election, the Great Reset conspiracy theory began trending globally on Twitter after being boosted by right-wing influencers in the United States and Europe. As our research demonstrates, the conspiracy continues to proliferate.

The narrative at the core of the Great Reset conspiracy is a dystopian fantasy of global control being imposed on the world’s population by the economic elite, including figures such as Bill Gates, who typically attend the WEF’s yearly conferences in Davos, Switzerland. According to this narrative, mRNA-based vaccines should be perceived as an instrument by which these elites can implement a single world government. One feature of the Great Reset conspiracy theory on Twitter is how it takes over hashtags and redefines them. Just as the WEF lost control over the messaging around their own initiative, so too has the

Figure 1: EARLY 2020 #MRNA CO-HASHTAG NETWORK IN TWITTER

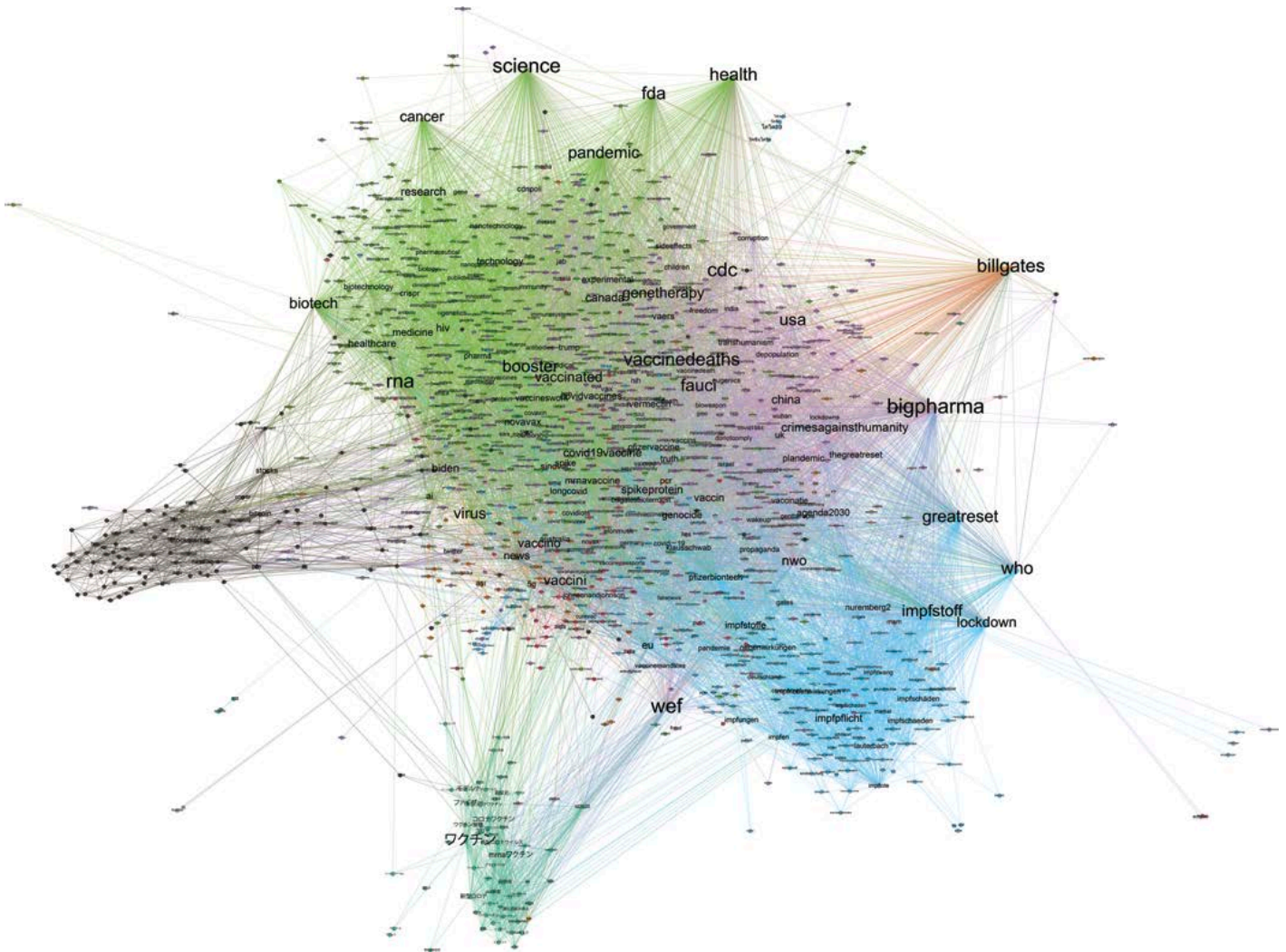


conspiracy narrative attached itself to and redefined numerous other discussions on the platform—notably, discussions around #mRNA.

By the end of 2022, the Great Reset conspiracy narrative had come to eclipse the scientific and economic associations of the #mRNA in our dataset. In Figure 3, you can also see that the purple community has many more references to frightening health concerns, most notably “vaccine death,” as well as references to the New World Order (#nwo) conspiracy theory, which can be traced back to far-right movements from the 1960s.

In addition to national communities, the visualization shows new communities, including a Japanese one in darker green expressing concerns about “depopulation” (#人口削減) and a Turkish one in orange expressing concerns about “vaccine victims” (#aşmağdurları). Around the perimeter you can also see large and highly connected terms shared by all of these communities, including *Great Reset*, *Bill Gates*, *WEF*, and *vaccine injuries*.

Figure 2: EARLY 2021 #MRNA CO-HASHTAG NETWORK IN TWITTER



The co-hashtag networks that have attached themselves to #mRNA in the last few years illustrate how scientists, scientific concepts, emerging technologies, and research institutions are appropriated and redefined as actors within the scripts of conspiracy theories. Notable examples include Bill Gates, who figures prominently alongside career scientist and policy advisor Anthony Fauci—both of whom are often presented as antagonists in the conspiracy narrative around mRNA.

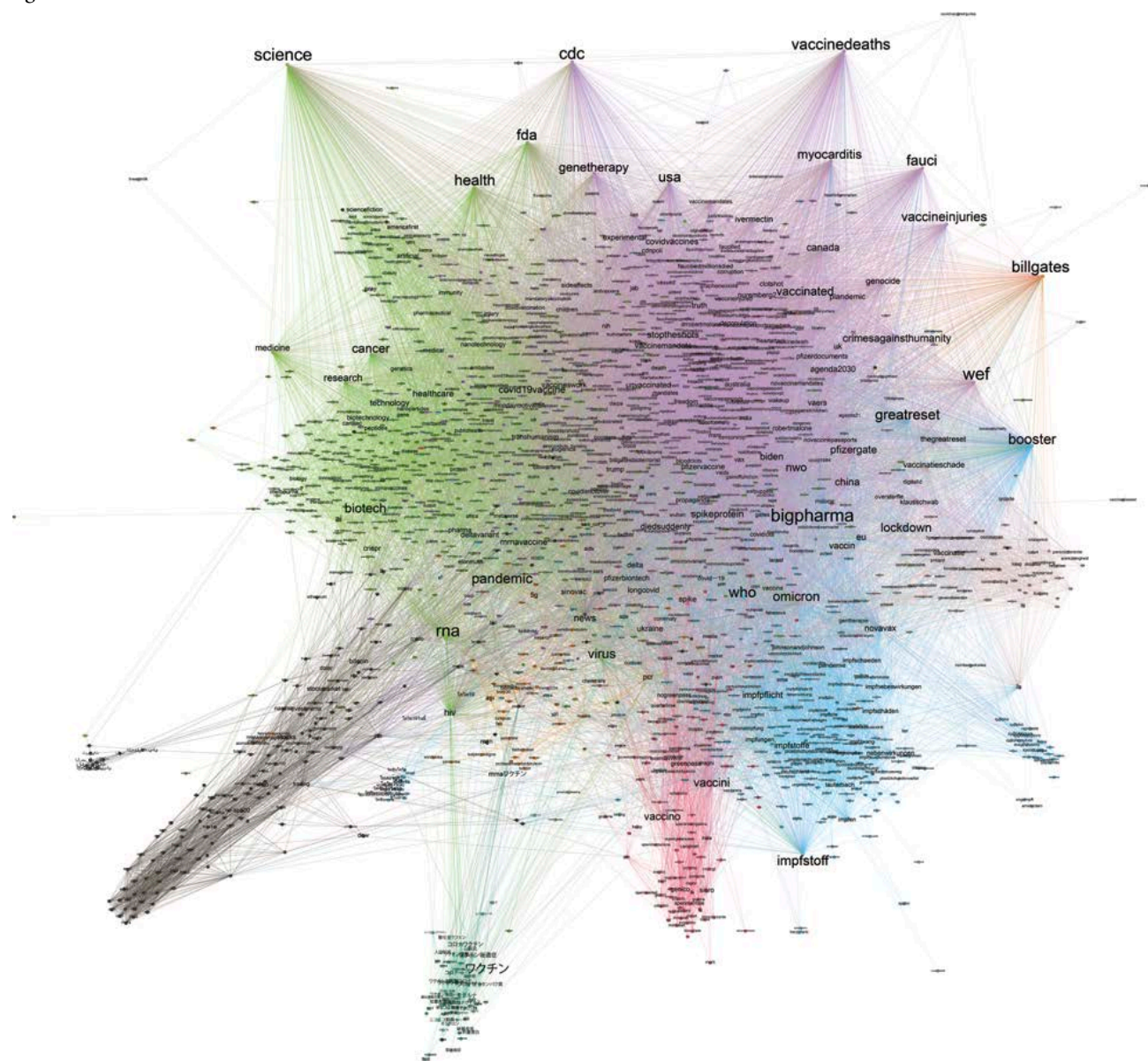
Evolving semantic associations like these may be understood as reflecting how the concept of mRNA became entangled with reactionary narratives that channeled fear and uncertainty about scientific innovations. Even more of this is evident today, in early 2023. With Twitter’s content moderation significantly enfeebled, #mRNA is merely one of many entry points into a vortex of conspiracies swirling around other science and technology hashtags, including #DigitalId and #SmartCities. Over time, these conversations become ever more semantically interconnected, mirroring a defining feature of conspiracy

theories themselves—that “everything is connected.”

Like many forms of narrative, conspiracy theories include networks of relations between characters, concepts, and other relevant entities like institutions or agencies. A common feature of conspiracy theories is the way that apparently unrelated concepts are pulled into seemingly coherent structures of associations. These associations may then combine into new, overarching narratives. During the pandemic, commentators began observing an intensified convergence of conspiracy narratives into what political scientist Michael Barkun refers to as “superconspiracies.”

Although the dynamic of narrative convergence is not unique to social media, there has been speculation that social media platforms may accelerate the process. For example, automated content recommendation on platforms like YouTube may draw users into rabbit holes of ever more extreme narratives. But differences between platforms also matter—disinformation and conspiracy theories evolve more rapidly on some platforms than

Figure 3: EARLY 2022 #MRNA CO-HASHTAG NETWORK IN TWITTER



All figures have been filtered for readability by removing the top 18 and bottom 100 nodes by weighted degree, with text sized by weighted degree, and color generated by modularity class. The complete dataset 2020–2023 contains 87,000 tweets.

on others. This can be attributed to design features of platforms that determine how things can be said, as well as to community standards, which are the terms and conditions governing what can be said.

On different social media platforms, users have a wide range of options—typically referred to as “affordances”—to actively connect narratives across communities and sources. For example, Twitter allows users to refer to multiple different entities in a single tweet, show them together in other images or videos accompanying the tweet, or tie them together through hyperlinks to external sources. Twitter and some other

platforms allow users to tag their posts with hashtags to make them more searchable. Social media is, in this sense, uniquely suited to enable the sort of interconnected narrative- and network-building that can be seen in conspiracy theories.

In our research, we’ve uncovered dense networks of associations where users have added multiple hashtags to their posts, a practice called “hashtag-stuffing.” We speculate that posters do this in the hopes of giving their posts higher degrees of visibility or retrievability. In extreme cases, which we have previously documented on Instagram, posts can include dozens of hashtags that are seemingly unrelated to each other or to

the actual content of the post they accompany. In this process, references to previously distinct narratives are brought together, intensifying the process of narrative convergence. Although the users' intention is probably to attract a slightly larger audience—and sometimes to make a quick buck—over time, we believe the effect may be to draw communities more tightly together, clustering around shared antagonism and antipathy.

This dynamic helps explain how conspiratorial keywords like #crimesagainsthumanity and #vaccinedeath glom onto seemingly neutral terms like #mRNA. What's more, this process of glomming seems to occur all over Twitter at the moment—so that if we started from a completely different and seemingly unrelated hashtag (#transhumanism or #FourthIndustrialRevolution, for example), we might have encountered much the same pattern, with those other hashtags becoming increasingly connected to the same “Great Reset” narrative emerging over time.

While this phenomenon of conspiracy narrative convergence appears very concerning, it does not necessarily reflect users' actual convictions. Twitter users do not necessarily take these narratives at face value. In fact, some researchers have claimed that paradoxically, unlike other major platforms, Twitter may actually have a “negative effect on conspiracy beliefs.” Though in some cases users may really believe in these narratives, in others it could be that they're merely “trying on” different ideas. Sociologist Aris Komporozos-Athanasίου refers to this use of social media as “speculative communities” built on an ambivalent form of belonging. Finally, because active Twitter users represent a relatively small section of the population, our findings should not be taken as somehow representing broad general tendencies.

What our methods do reveal is that Twitter's affordances appear to enable the building and spread of conspiratorial narratives and their insinuation into discussions on all kinds of seemingly unrelated topics. This is a significant matter of concern when it comes to public understanding of science and technology: the unchecked spread of these conspiratorial narratives could severely undermine the networks of trust on which society is built.

The mechanisms that we have uncovered here have implications for policies and strategies that combat misinformation and disinformation, which currently do not take the convergence of social media affordances and the narrative structure of conspiracy theories into account. The first step (as in many misinformation-countering strategies) is to consider changing platform governance. In addition to different affordances, platforms have various guidelines that set acceptable speech and define rules on what is considered illegal, harmful, or antagonizing content. As of March 2023, Twitter's guidelines aim to restrict “violence, harassment, and other similar types of behavior,” as these might “discourage people from expressing themselves.” The platform's rules thereby address, among other things, violence, terrorism, violent extremism, child sexual

exploitation, abuse, harassment, hateful conduct, graphic violence, and adult content. But these rules do not address hashtag stuffing or the process that enables disparate ideas and narratives to be easily glommed together in formats that make them more easily searched and shared.

The standard mechanism through which platform rules are enforced is content moderation. Decisions concerning what content is allowed or removed on a platform are a continually negotiated balance among free expression, illegality, harm, and toxicity. On mainstream social media platforms, restrictions on disinformation are often broader than legally necessary (at least in liberal democracies), but controls are typically implemented in an uncoordinated and reactive manner. As a result, emerging alternative platforms such as Telegram, Gab, and TruthSocial, through their promises of lax content moderation, might exert a high degree of attraction for antagonistic actors and narratives that have been deplatformed elsewhere. Whether content moderation ultimately leads to a game of whack-a-mole as users move from platform to platform is an ongoing subject of discussion among researchers in this field.

Our glimpse into the network of associations for the hashtag #mRNA on Twitter reveals the complex dynamics through which science becomes the object of disinformation. Although scientists and others have interpreted these patterns as speaking to the contested nature of science in these networks, our work indicates that social media can create resonances between otherwise disparate issues and that scientific concepts might be roped into conspiracy theories and disinformation—along with everything else. These complex networks of associations suggest that simplistic approaches to combating such conspiracy theories—either by moderating content on social platforms or educating users to be skeptical of information—may not be effective.

This does not mean that there is no way out of the rabbit hole. There are many strategies that have not been tried. Designing algorithms that encourage consensus as opposed to exacerbating division could help, as would programs and policies that build bridges among users rather than dividing them along ideological or party lines. And although it won't result in instant change, continuing to invest in public education that explains scientific methods—what scientists know but also what they do not know—could rebuild trust in science as a whole and make scientific concepts less likely to be glommed in with other conspiracies. Ultimately, to stall conspiracies online, researchers will need to look to the real world and acknowledge and address why antagonism and antipathy toward science become attractive.

Marc Tuters is an assistant professor in the media studies faculty at the University of Amsterdam. Tom Willaert is a postdoctoral researcher at the Brussels School of Governance and the Studies in Media, Innovation, and Technology research center at Vrije Universiteit Brussel in Belgium. Trisha Meyer is an assistant professor in digital governance and participation at the Brussels School of Governance at Vrije Universiteit Brussel in Belgium.