

Socrates Untenured

The Social Media Propaganda Problem

Is Worse Than You Think

In 2016, the Internet Research Agency (IRA), the organization responsible for much of the Russian state-sponsored social media campaign documented in the Mueller report, created a Facebook group called “Being Patriotic.” In September of that year, the group posted an image of a weathered veteran prompting readers to “Like & share if you think our veterans must get benefits before refugees.” The caption claimed that “liberals” wanted to invite 620,000 refugees across the US/Mexico border while over 50,000 homeless veterans were “dying in the streets.”

The claim about refugees came from one of then presidential candidate Donald Trump’s stump speeches; it has since been refuted by Politifact. Nonetheless, the meme was shared by more than 640,000 Facebook users. This is just one example of how the IRA worked to undermine democratic functioning in the United States and throughout Europe. Its campaign targeted social media users on both the right and the left by producing and sharing misleading and intentionally divisive images and other content—up to and including manufacturing and selling “Black Matters” (sic) T-shirts and “LGBT-positive” sex toys.

The scope and audacity of the Russian influence campaign, both in the lead-up to the 2016 election and since, have revealed startling and unanticipated ways in which new technology, particularly social media sites such as Facebook and Twitter and apps such as Whatsapp that permit peer-to-peer

dissemination of content, have made society more vulnerable to disinformation. These emerging vulnerabilities demand a policy response. But their distinctive character also creates new challenges. Social media propaganda is different from other forms of propaganda, and this matters for how it should be regulated.

As both scholars and policy wonks are keen to emphasize, there is a distinction between disinformation (content purposefully shaped to mislead, usually for political or economic purposes) and misinformation (false or misleading material that is shared without deceitful purpose). That we should seek effective policy responses to disinformation, particularly when generated as part of an influence campaign waged by a hostile foreign power, seems obvious. But regulating misinformation is another matter. The right to be wrong is a central tenet of the liberal tradition, without which society could not have freedom of thought or freedom of speech. That some political agendas happen to be served by widespread misconceptions or errors is a matter of luck, and not something that can be remedied by policy changes.

The attitude we have just sketched, which relies on the distinction between misinformation and disinformation, made sense in a pre-Twitter media environment. But reflection on how beliefs spread in social networks, and particularly in online social media platforms, suggests that the distinction between disinformation and misinformation is no longer

tenable—and worse, is actively misleading.

Human learning and knowledge are deeply social. People tend to adopt their beliefs from the testimony of others. How, for instance, do you know that tomatoes are safe to eat? Or that you should not inhale asbestos? Why do you believe whatever you believe about genetically modified foods? It is unlikely that you have done any tests on the safety of these things yourself. Instead, you learned these things, and most of the other facts you know, from peers, parents, teachers, books, articles, and websites.

This type of social learning is essential to the development of complex human cultures, and makes possible a range of defining human activities such as economic markets and technological innovation. However, as we discuss in our book *The Misinformation Age*, this ability comes with a downside: it opens the door to the spread of false beliefs.

The IRA, as well as other interest groups that use social media to shape people's beliefs, is savvy about the social aspects of belief. The organization takes advantage of these social features, and the way it does so blurs the line between misinformation and disinformation.

Think about the veteran meme described above. It was created by Russian agents with the intent to mislead, making it a piece of disinformation in the classic sense. However, its vast reach was the result of sharing by true believers, with no intent to mislead. In this sense, it was also a piece of misinformation—false information shared without deceitful intent. At the same time, the IRA purposefully shaped the meme to go viral: its intent was to create disinformation that would quickly transform into misinformation. In this way, the group's message was amplified in a way it never could have been had it been disseminated directly.

Of course, even before social media, friends could pass along falsehoods they had picked up from sources of disinformation. But there is a key difference. In pre-social-media information environments, only a small number of organizations had the capacity to broadcast content to large audiences. But on social media, memes, images, and claims can be widely rebroadcast by users. The result is that peer-to-peer transmission—that is, propaganda in the guise of misinformed individual speech, as opposed to disinformation—plays a much more significant role in how ideas are spread.

There are two reasons the strategy of creating disinformation that becomes misinformation is so powerful. First, people trust their friends and others whom they perceive to be like themselves. In deciding what pieces of information to believe, people take the perceived trustworthiness of the sharer to be very important. We more readily accept information shared by a friend or trusted peer than from someone we do not know.

Second, when disinformation transforms into misinformation, it is harder to detect. When people see memes shared by friends, they do not usually know where they originated. If all of Russia's disinformation were readily traceable to a single source, it would be easy to learn to discount that source. This is not possible when memes are spread from peer to peer. In this way, the structure of new media creates opportunities for propagandists, because the viral spread of misinformation obscures the role of bad actors in its creation.

So, is the veteran meme misinformation or disinformation? It is not clear that either designation fits. Calling it disinformation downplays the role of social connections and social trust in its propagation. Calling it misinformation downplays the intent to mislead and the ways it is purposefully shaped to become effective misinformation. Once we understand these processes, the distinction becomes blurred, and we can see that better language is called for.

There is a further sense in which both terms—misinformation and disinformation—can be misleading. We often think of online propaganda as equivalent to spreading falsehoods. The caption of the veteran meme, for example, includes a false claim about “liberal” plans. This is a limited view, however.

For instance, “Being Patriotic” and other IRA-instigated Facebook groups posted lots of content intended solely to connect with their members. On Instagram the IRA account @blackstagram posted an uplifting image of women's legs in different skin tones and the line, “All the tones are nude! Get over it!” (An animal lovers' group on Facebook posted cute animal memes, and an LGBT group shared images from a coloring book of a heavily muscled Bernie Sanders character called “Buff Bernie.”) These efforts did not mislead, but instead built trust with readers as a basis for later manipulation, for instance by encouraging black voters to boycott the election or by encouraging Bernie supporters not to vote for Hillary Clinton.

Furthermore, online propaganda is often aimed at directly promoting some action, without promoting some false belief state. You might consider the difference between the statements, “candidate X wants to admit 620,000 refugees into the US” and “f*** the elections!” It's clear how the first statement might hurt a political candidate, but the second statement might do the same with no falsehood involved. This was the sort of propaganda that Russia used to try to discourage African Americans from voting in the 2016 election.

And true or nearly true facts can be misleading when shared in the wrong ways. One meme, which appeared on Twitter during fall 2018, contained these statements:

Reflection on how beliefs spread in social networks suggests that the distinction between disinformation and misinformation is no longer tenable—and worse, is actively misleading.

“449,000 Californians received a jury summons last year to which they replied ‘I am not a citizen, therefore I cannot sit on a jury.’ The number one source for jury summons candidates is the voter registration list. Think about it!”

In a case like this, the statements can all be true. (As it happens, the first claim apparently originated with Mark Meuser, a Republican candidate for California secretary of state, in an interview with the *Santa Clarita Gazette*. It is unclear whether the figure is correct. The second claim is not strictly true; Department of Motor Vehicles records are the primary source of jury summons candidates for most California Superior Courts. But voter rolls are also used.) And yet the obvious conclusion—that voter fraud is rampant in California—can be false. It might be wrong to call this misinformation or disinformation if all the information is true. But the hoped-for result is nonetheless a false belief on the part of the reader.

Given the wide variety of forms that online influence campaigns take on, it might be better to group them under the heading of “propaganda” than to use either “misinformation” or “disinformation” to describe them. Propaganda implies an intent to shape the minds and behaviors of a population, but not necessarily through falsehood. There are many kinds of propaganda, and the examples discussed here—of disinformation that becomes misinformation, of direct prompts to action, of true statements shared in misleading ways, of ill-intentioned trust building—all fall under this heading.

Once we recognize that both disinformation in the classical sense and the misinformation it rapidly becomes can be deliberate aspects of a propaganda campaign, we can see that policy approaches that focus only on disinformation will be ineffective. Cutting out the source of false beliefs will not stop them from persisting and spreading once those beliefs have taken hold. And the time and effort needed to create and distribute effective memes is miniscule compared with the amplification they can receive when spread in a social network. On the other hand, policies that censure the individuals who unwittingly share or promote propaganda are untenable on free speech grounds.

One way to cut this Gordian knot is to recognize that misinformation spreads on social media not only because users elect to share it when it crosses their screens but also because divisive and emotional content, which elicits reactions at an elevated rate, tends to be amplified by the recommendation algorithms used by social media sites. Just as modern propagandists are increasingly sophisticated about what content is most likely to elicit reactions and be shared, so too are social media companies incentivized to find and promote such content to maximize the time users spend on the site.

In other words, social media sites already actively curate what content their users see, in what order, and with what context. But there is no transparency in how this curation is done and virtually no rules to protect users from malicious actors. It is here that a policy intervention is most attractive. Although individuals may have a right to be misinformed and to share their false beliefs with others, there is no legal framework entitling them to have those beliefs amplified by algorithms. And there is a clear public interest in ensuring that the profit motive of the platforms on which most social interactions now take place does not interfere with the effective functioning of the nation’s democracy.

Social media companies must be held responsible for designing promotion and recommendation algorithms that are sensitive to the difference between genuine, user-generated, emotive content and disinformation designed for peer-to-peer sharing. Content that has been identified, either by human editors or machine learning methods, as likely propaganda should not be promoted on users’ personal feeds—even when it has been shared or liked by their friends or those whom they follow.

Also needed is a regulatory framework for reviewing these algorithms and for identifying when new propaganda strategies emerge that exploit previously unrecognized vulnerabilities. Most important, any policy solution to these problems needs to recognize that online propaganda is, in the words of the philosopher of science Bennett Holman, an asymmetric arms race. Purveyors of disinformation are constantly evolving their methods to exploit current systems, and social media firms must be able to adapt quickly to emerging threats.

Cailin O’Connor is an associate professor of logic and philosophy of science, and director of the Philosophy, Political Science, and Economics program, at the University of California, Irvine. James Owen Weatherall is a professor of logic and philosophy of science at UC Irvine, and author of The Physics of Wall Street. They are coauthors of The Misinformation Age: How False Beliefs Spread.