We made deceptive robots to see why fake news spreads, and found a weakness

November 29 2018, by Carlo Kopp

Credit: Andrea Piacquadio from Pexels

Only a small amount of fake news is needed to disrupt any debate or

discussion on an issue, according to research published today in *PLOS ONE*.

But there is a way to discourage those spreading fake news, and even to wipe it out altogether.

The research is experimental, based on modelling and simulations, but it does at least show that it is possible to counter the spread of misinformation.

The rise of fake news

The spread of mischievous and false information has plagued human societies for centuries.

In this era of instantaneous global digital connectivity, the current incarnation of "fake news" has become a scourge and is exploited for personal or political gain.

Social media, designed to encourage users to contribute and share content, has become the great enabler of the spread of fake news.

From nations meddling in the politics of democracies and political parties trying to manipulate public opinion, to a profit-centred "fake news" industry, all have exploited this spread for gain, sowing confusion and discord in the victim populations they target.

The simulation game

We did some experiments aiming to understand the more fundamental mechanisms determining the behaviour of fake news in populations.

We were especially interested in two questions:

1. how much impact fake news can have on consensus-forming in a population
2. the impact of the cost of distributing fake news on its ability to infest a population.

In the real world, costs can be external, such as fines, penalties, exclusions, expenditures in creating and distributing fakes; or they can be internal, such as feelings of loss or embarrassment due to being ridiculed or shamed.

The tool we used was an evolutionary simulation, in which simple software robots in a population interact, playing the well-known Prisoner's Dilemma game. Basically, a prisoner who betrays another wins big, while the betrayed loses badly, while both only win modestly if they cooperate, and suffer equally if they betray one another.

Unlike previous work in this area, we made some of these software robots a little devious, by adding code that allowed them to deceive each other. The victim of such a deception is made to be confused about the opposing player's intent, or convinced the opposing player is an unselfishly cooperative "good guy".

Our code made use of our work in information-theoretic modelling of deceptions, allowing known deceptions to be mapped into game theory models. Every deceiver in the simulation incurred a cost when they deceived, which was then subtracted from the payoff they earned in the prisoner's dilemma game.

How much fake news to disrupt consensus?

We found that even a very small percentage of deceiving players in the

population – in our simulations less than 1% – could catastrophically disrupt cooperative behaviours in the simulated population.

In the extreme case of cost-free deceptions – where fake news producers are unhindered – cooperative behaviours vanished altogether. Only where the cost of deceptions was larger than zero, did cooperation survive. Where costs were very high, cooperation actually thrives.

We also found that for all simulations, the ability of deceiving players to survive depended very strongly on the cost of deceptions. If the cost was high enough, deceivers could not survive in the population.

Applying this to the spreading of fake news, very high costs will lead to its extinction.

From experiment to real world

What do these experimental results tell us about the real world of fake news distribution in social and mass media?

The first and arguably more important result is that very little fake news is required to create mayhem in a population, and prevent consensus forming that is critical to public debates. Whether victims are confused, or believe falsehoods, is immaterial. It's their ability to reach consensus that is disrupted.

Our modelling focused on small groups of influencers who actively debate issues. Where influencers cannot agree, followers in turn cannot align to a consensus. This is one of the reasons why fake news is so destructive to democratic societies.

The second result of broader interest is that attaching a high cost to the production, but especially the distribution of fake news may prove to be

the most effective tool we have to defeat its spread. A high societal investment in raising these costs is worthwhile, because the effects of fake news are so disruptive.

Breaking the chain

Information warfare research over a decade ago found that proxy delivery was a major multiplier in the distribution of toxic propaganda.

For example, mass media distributing violent imagery and footage produced by terrorists were acting as proxies for the terrorists producing the propaganda, whether they knew it or not.

Social media users who share fake news are likewise acting as proxies for the producers of fake news. Such users are typically cast as victims of fake news – which they usually are – but every time they share fake news they become participants in the fake news producer's deception.

Attaching a cost to the distribution of fake news in social media is not straightforward. The informal outing of habitual posters of fake news is one option, which accords with the evolutionary psychology of cheater detection.

Social media organisations such as Facebook say they are trying to be more proactive in detecting fake news and false news either by machine learning technology or third-party fact checkers, and says it has had some recent successes.

But both of these ideas run into the stickier problem of determining exactly what is or is not fake news. Unpalatable facts are too often labelled as "fake news".

Fact checkers' reliability and objectivity can vary widely – ground truths

are often obscured by bias, and limitations in understanding.

At this time, contrary to claims by some social media providers, AI is not up to the task of finding and weeding out fake news, which puts the onus back on us humans.

We can all help simply by thinking a little before we like, share or retweet any information on social media. Maybe do a few search checks to see if the information is known to be true or fake.

Pest control is an established practice in biological ecosystems, and is clearly overdue for the information ecosystem.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Provided by The Conversation