Nature
24 July 2019

Disinformation’s spread: bots, trolls and all of us

By Kate Starbird

. . .Much ink has already been spilt on the algorithms, business models and human impulses that make the social-media ecosystem vulnerable to disinformation, the purposeful spreading of misleading content. Although the major tech companies are often insufficiently open about, or motivated to fix, the problem, they have begun to take action against what Facebook calls “coordinated inauthentic behavior”.

But disinformation is not as cut-and-dried as most people assume: those behind disinformation campaigns purposely entangle orchestrated action with organic activity. Audiences become willing but unwitting collaborators, helping to achieve campaigners’ goals. This complicates efforts to defend online spaces.

When my lab studied the online activism around #BlackLivesMatter, the conspiracy theories that crop up after crises, and the Syrian conflict, we uncovered disinformation campaigns promoting multiple, often conflicting, views. At first, I overlooked their influence, hoping to get to more-important underlying phenomena. But eventually I began to see how disinformation networks were warping online conversations and global political discourse, and I changed my research focus. Years later, the sorts of misconceptions that had led me to discount disinformation continue to hamper responses to the threat.

Perhaps the most common misconception is that disinformation is simply false information. If it were, platforms could simply add ‘true’ and ‘false’ labels, a tactic that has often been suggested. But disinformation often layers true information with false — an accurate fact set in misleading context, a real photograph purposely mislabelled. The key is not to determine the truth of a specific post or tweet, but to understand how it fits into a larger disinformation campaign.

Another misconception is that disinformation stems mainly from agents producing false content (paid ‘trolls’) and automated accounts (‘bots’) that promote it. But effective disinformation campaigns involve diverse participants; they might even include a majority of ‘unwitting agents’ who are unaware of their role, but who amplify and embellish messages that polarize communities and sow doubt about science, mainstream journalism and Western governments.

This strategy goes back decades. It was laid out most explicitly by Lawrence Martin-Bittman, who defected from Czechoslovakia to the West in 1968 and became a prominent academic (L. Bittman *The KGB and Soviet Disinformation*; 1985). Historically, manipulating journalists was a primary strategy. Now, social-media platforms have given voice to new influencers and expanded the range of targets. We see authentic members of online communities become active contributors in disinformation campaigns, co-creating frames and narratives. One-way messages from deliberate actors would be relatively easy to identify and defuse. Recognizing the role of unwitting crowds is a persistent challenge for researchers and platform designers. So is deciding how to respond.

Perhaps the most confusing misconception is that the message of a campaign is the same as its goals. On a tactical level, disinformation campaigns do have specific aims — spreading conspiracy theories claiming that the FBI staged a mass-shooting event, say, or discouraging African Americans from voting in 2016. Often, however, the specific message does not matter. I and others think that the pervasive use of disinformation is undermining democratic processes by fostering doubt and destabilizing the common ground that democratic societies require.

Perhaps the most dangerous misconception is that disinformation targets only the unsavvy or uneducated, that it works only on 'others'. Disinformation often specifically uses the rhetoric and techniques of critical thinking to foster nihilistic scepticism. My student Ahmer Arif has compared it to listening to static through headphones. It is designed to overwhelm our capacity to make sense of information, to push us into thinking that the healthiest response is to disengage. And we may have trouble seeing the problem when content aligns with our political identities.

Disinformation campaigns attack us where we are most vulnerable, at the heart of our value systems, around societal values such as freedom of speech and the goals of social-media platforms such as 'bringing people together'. As individuals, we need to reflect more on how we interact with information online, and consider that efforts to manipulate us may well be coming from within our own communities.

Before social-media platforms can tackle how to identify and combat disinformation, they need to work out which behaviours are problematic, even when such behaviours might be good for profits. And they need to acknowledge that technology is not neutral, that their platforms embed certain values. If supporting democratic discourse is one of those values, then companies need to own that, to anchor their responses in that value, and not be cowed by disingenuous claims of bias from those who seek to benefit from the continued spread of disinformation.

As researchers and policymakers, we have to go beyond trying to measure the impact of individual disinformation campaigns using simple models of inputs (for example, messages posted by bots or trolls) and outputs (such as likes, retweets or even votes). We need models that can encompass how disinformation changes hearts, minds, networks and actions. Solving this will take a level of collaboration across platform designers, policymakers, researchers, technologists and business developers that is, frankly, hard to imagine. A free society depends on our finding a way.

Nature **571**, 449 (2019)

doi: 10.1038/d41586-019-02235-x

(Kate Starbird studies human–computer interaction and crisis informatics at the University of Washington in Seattle. Previously she played for ten years in the WNBA before earning her PhD at the University of Colorado at Boulder)
(accessed on 7.31.19 at https://www.nature.com/articles/d41586-019-02235-x)