Devil in the algorithm

Social media platforms are damaging democracy, and it's not primarily about what speech they do or don't moderate.

From the Editors in the June 1, 2022 issue

(Photo by akshar dave on Unsplash)

Elon Musk's acquisition of Twitter has left people anxious about the social media platform's future. Will it become a cesspool of disinformation and hate? Will Donald Trump come back? Will Musk's emphasis on uninhibited speech create an environment that inhibits meaningful conversation?

These are important questions. But when it comes to social media's impact on civic life, the core issue is not free speech versus moderation. "Free speech is fundamentally about neutrality with regard to content," writes journalist Matthew Yglesias in his newsletter—and "Twitter is not a neutral platform." The problem is the algorithm that determines which public posts users see on Twitter (or their Facebook news feed) in the first place. An engagement algorithm is biased toward whatever motivates people to do ever more posting, replying, liking, and sharing.

And it's become clear that such algorithms have done a great deal to erode the reliability of public information and the norms of civil discourse.

The algorithms aren't designed to promote high-quality conversation, of course. They're designed to maximize profits [for the powerful people who control them](#). The platforms tend to insist that their algorithms also serve users well, though they don't offer much transparency as to how.

What is clear is that engagement algorithms reward hyperbole, vitriol, and conflict. They turn conversation into a contest you can win, a seductive and destructive shift. They even reward falsehood itself. Yglesias points out that a false news story—intentional or not, salacious or not—performs better than any of 99 competing true ones simply because it's distinctive. So the algorithm keeps showing it to more and more people.

Writing in [*National Affairs*, Jon Askonas and Ari Schulman](#) explore a subtler problem: engagement algorithms work against any sense of genuine community. They decide what you see in your feed; you and your friends do not. But the very notion of open discourse, whether moderated or wholly uncensored, is only legible within the context of a polis—of a political community, however diverse, with some measure of mutual experience and shared reality. Such a context provides both the oxygen and the guardrails for genuine, meaningful speech to flourish.

Engagement algorithms provide something very different: a structural expression of some of people's worst impulses. Individuals are responsible for their own behavior, but the algorithm prods and incentivizes them to be their worst selves—by design, because that's what most engages others. Christians have a word for this sort of problem: sin. The algorithm might not be what most of us want on social media, but it does reflect what many of us do there. This gap between what we want and what we do is central to Paul's understanding of sin's power over our lives, a power that is far bigger than individual willpower (Rom. 7:15–20).

To be sure, there is benign and even virtuous activity on social media as well. But the harm social media is doing to democracy is not tangential to the platforms' design; it's intrinsic to it. It won't get better without a fundamental change to the kinds of behavior they encourage and reward.