Shorenstein Center on Media, Politics and Public Policy

March 2018

Transparency: What's Gone Wrong with Social Media and What Can We Do About It?

By Wael Ghonim, Entrepreneurship Fellow, Fall 2017, and author of *Revolution 2.0*, and Jake Rashbass, Knox Fellow and Master in Public Policy student, Harvard Kennedy School

---

HARVARD Kennedy School
SHORENSTEIN CENTER
on Media, Politics and Public Policy

Table of Contents

1. Introduction: Social Media between Utopia and Dystopia 3
2. Optimizing for Engagement: The Structural Problem of Social Media Platforms 3
3. Getting Political 5
4. The Personalization Myth: What a User Wants 6
5. Short-Term Business Interest vs. Long-Term Public Interest 7
6. Transparency: A Strategic Framework for Countering Mobocracy 8
7. Towards A Solution: Public Interest APIs 9
8. Areas for Further Research 10
9. Endnotes 11

Introduction: Social Media between Utopia and Dystopia

“The Internet is the largest experimentation involving anarchy in history. […] Never before in history have so many people from so many places had so much power at their fingertips. And while this is hardly the first technology revolution in our history, it is the first that will make it possible for almost everybody to own, develop and disseminate real-time content without having to rely on intermediaries.”

— Eric Schmidt and Jared Cohen[1]

“We're very different than a media company. At our heart we’re a tech company. We hire engineers. We don’t hire reporters. No one is a journalist. We don’t cover the news. But when we say that, we’re not saying we don’t have a responsibility. In fact we’re a new kind of platform… [and] as our size grows, we think we have more responsibility.”

— Sheryl Sandberg, Chief Operating Officer at Facebook, 2017[2]

When social media platforms like Facebook, Twitter, and YouTube emerged, they were heralded by scientists, activists, and artists alike as an unprecedented tool to advance human civilization. By democratizing the publication and distribution of content, they would disrupt age-old power structures. No longer would the rich and powerful have a monopoly on public attention. Power would reside with the people.

That perspective has changed dramatically. Power has been centralized in the hands of an oligopoly of platforms. Barely a day goes by without another exposé on the adverse effects of social media—empowerment of extreme groups, hardening of echo chambers, dissemination of polarizing disinformation, fostering of emotional and mental instability.

The pathologies of platforms are largely a consequence of algorithms—self-learning predictive mathematical models that optimize primarily for user engagement. But there’s reason for hope. This paper attempts to trace what went wrong with social media and proposes an initial step to reverse course and put platforms on track to being a productive, responsible, and ethical force.

Optimizing for Engagement: The Structural Problem of Social Media Platforms

Algorithms pervade modern life. They impact our financial options, our shopping choices, our social interactions, and our access to information. On a basic level, an algorithm is a program that decides based on a set of ranking criteria which option from a set of alternatives to prioritize. Because of the explosion of content on social media platforms, with millions of posts uploaded each minute, algorithms determine the content that will be presented to the individual user.

Platforms optimize their algorithms with one goal in mind: getting and holding users' attention. User attention is a platform's ultimate source of revenue. By keeping users engaged, platforms are able to capture an audience that can be sold things and ideas.

Because people's attention is a scarce resource, social media algorithms are highly personalized to optimize for each user's engagement. This means they will rank highest the content that each user is most likely to interact with (using actions such as clicks, likes, comments, and shares as indicators of preference). To determine the content users find most engaging, platforms collect and use hundreds and sometimes thousands of what are known as "signals" to predict whether a user will engage with a given piece of content. These signals match users with content. For the user, these signals might include their age, location, gender, device type, previous engagement habits, and friends' engagement habits. For the content, these signals might include its medium (text, video, or image), length, time of publishing, and other users' engagement with it (i.e., what share of the users already presented with the content have engaged with it). The origins and values behind that content are largely irrelevant—what matters most is that the content is popular and that the user remains engaged.

Optimizing for user engagement entices users to spend increasing lengths of time on a platform, which researchers have shown can have negative consequences on users' well-being.[3] But for the platform, more time means a greater opportunity to advertise and earn revenue. Moreover, because platforms have personalized data on each user's behavior and preferences, advertisements can be micro-targeted to maximize the probability of a user clicking on an ad.

This business model works. As platforms' user base has increased, advertising revenues have seen staggering growth.[4] The amount of time that users spend on platforms has reached great heights, too. Mark Zuckerberg, Facebook's founder and Chief Executive Officer, announced in 2016 that the average user was spending 50 minutes a day on the platform.[5]

In the early days of social media, the algorithms curated the purely social and broadly innocuous content that dominated the platforms, such as family and friends' updates and pet photos. Because this material was largely apolitical, there was little need to evaluate the nature of content or the credibility of its author. The algorithms could safely distribute materials based on engagement statistics alone.

Today, however, social media are a hotbed of political activity. Barack Obama's 2008 presidential campaign is often identified as the birth of social media as a political tool in the United States.[6] But the change is by no means limited to the U.S. Social media played a significant role in the so-called Arab Spring protests of 2010-2011, particularly in Egypt.[7] Social media has also provided a platform for various social movements.[8] At the same time, platforms have become a major source of news for many users, with a 2016 Pew survey finding that over 60 percent of adults in the U.S use social media platforms to find out about current affairs.[9]

Getting Political

When applied to politics and social issues, algorithms that prioritize content that engages users regardless of its authenticity or credibility can breed mobocracy—a situation where mob behavior is incentivized and empowered. There are at least three reasons why this can happen.

First, optimizing for engagement fosters polarization and sensationalism. In a study of 200,000 press releases and Facebook posts, Pew Research Center found that U.S. members of Congress received 50 percent more "likes," three times as many comments, and twice the number of shares for posts that expressed "indignant disagreement" than for those that expressed bipartisan sentiments (see fig. 1).[10] On Twitter, analysts observed a drop in users' engagement with President Donald Trump's tweets during his first 100 days in office as the content became more "reserved." Pre-election engagement levels were observed only for tweets on controversial subjects, some of which were composed only of capital letters.[11]

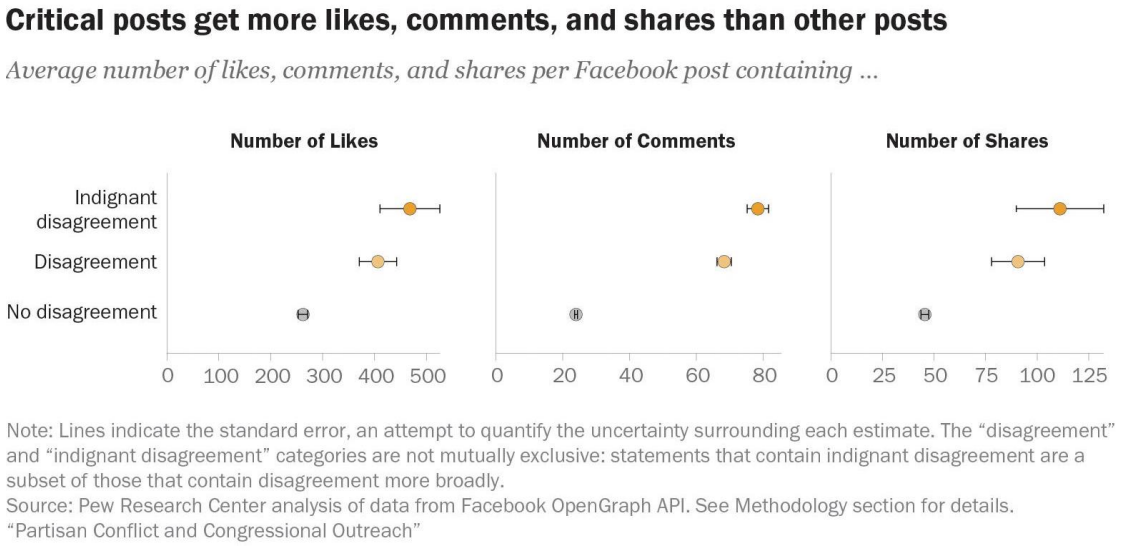

Fig. 1 (Source: "Partisan Conflict and Congressional Outreach." Pew Research Center, 23 Feb. 2017, www.people-press.org/2017/02/23/partisan-conflict-and-congressional-outreach/)

Second, platform algorithms have led to an explosion of misinformation and fake news. By prioritizing engagement without regard for the accuracy of content, users engage with content without any indication of its truthfulness. The 2016 U.S. presidential election revealed the extent of the problem. Take, for example, Jonathan Albright's research findings on Russian-created Facebook pages targeting American citizens. These pages, with names such as "Blacktivists", "United Muslims of America," and "Secured Borders" fraudulently presented politically sensitive issues and social groups and issues. They collectively gained over 340 million engagements before being deleted by Facebook.[12]

Third, platform algorithms have led to the prevalence of filter bubbles. A 2015 study of over 10 million Facebook users' friend networks showed a distinct ideological segregation. The median proportion of self-identified conservatives within self-identified liberals' networks was one in five. Liberals accounted for even a smaller share of conservatives' networks.[13] Social media platforms have morphed into the digital equivalent of the ideological demographic segregation that journalist Bill Bishop described as the "Clustering of Like-Minded America."[14] As Hunt Allcott and Matthew Gentzkow have shown, ideological segregation contributes to the spread of fake news and misinformation because "people who get news from Facebook (or other social media) are less likely to receive evidence about the true state of the world that would counter an ideologically aligned but false story."[15] Moreover, as technologist and activist Eli Pariser has shown, filter bubbles foster confirmation biases because "consuming information that conforms to our ideas of the world is easy and pleasurable [whereas] consuming information that challenges us to think in new ways or question our assumptions is frustrating and difficult. [...] As a result, an information environment built on click signals will favor content that supports our existing notions about the world over content that challenges them."[16]

The situation is likely to get worse. Over time, as a platform gathers more and more data on its users, and as the technology it uses to analyze their engagement improves, the ability to personalize the individual's engagement profile will increase.

The Personalization Myth: What a User Wants

"We shape our buildings, and afterwards our buildings shape us."
— Former British Prime Minister, Winston Churchill[17]

One might argue that by optimizing for engagement, platforms' algorithms are merely presenting users with the content they prefer to see. If a user engages with a piece of content, surely that is the content that the user prefers. If an idea is good and worthy, it will be distributed and perhaps even go viral. If it's bad, users won't engage with it and it will be curbed out of the network. According to this logic, optimizing for engagement is the most efficient and neutral means by which to distribute ideas.

The argument is flawed. Daniel Kahneman, who won a Nobel Prize for his work on behavioral economics, describes two systems of human cognition. System One is for fast and instinctive thought, requiring minimal mental exertion or concentration, which makes the thinking process highly efficient, but comes at the expense of errors that could be avoided by further reflection. System Two represents the reverse: a slower means of thinking that demands significantly more engagement and effort, but avoids the instinctive errors that come with System One.[18] So, for example, offering someone junk food would trigger System One thinking. It would produce an instinctive reaction, resulting in either the quick acceptance or quick rejection. This response differs from

the more deliberated System Two thinking that would be triggered by asking someone to plan his or her food diet for a year.

Engagement-driven algorithms prey on System One thinking. The content that a distribution algorithm identifies as engaging is likely to attract the user's interest regardless of how, upon reflection, the user might generally want to use the platform. Distribution algorithms aim to keep the user on the platform for as long as possible, which is not necessarily how long the user, upon reflection, would want to use it. In other words, there is a difference between providing a user what they reflectively want and what they can be tempted to consume.

danah boyd, president of the Data & Society Research Institute and a principal researcher at Microsoft, was one of the first to spot the dangerous potential of an internet driven by the attention economy. In a speech delivered at the Web 2.0 Expo in 2009 she warned: "Our bodies are programmed to consume fat and sugars because they're rare in nature.... In the same way, we're biologically programmed to be attentive to things that stimulate: content that is gross, violent, or sexual and that gossip which is humiliating, embarrassing, or offensive. If we're not careful, we're going to develop the psychological equivalent of obesity. We'll find ourselves consuming content that is least beneficial for ourselves or society as a whole."[19]

Short-Term Business Interest vs. Long-Term Public Interest

Platforms' business models give them little incentive to change what they're doing. As noted earlier, optimizing for engagement means more opportunities to show ads. In addition to that, it's very challenging to identify and measure what's in the public interest. Because of this, deviating from maximizing engagement comes with a decrease in company revenue and an inability to measure success. Even the most idealistic engineers tasked with tweaking the algorithms lack an incentive to identify and prioritize societal well-being. Their performance reviews are largely tied to the value they add to the company's revenue and user base.

This tension in algorithm development between short-term shareholder value and long-term public interest is hardly limited to social media platforms. Mathematician Cathy O'Neil, who was tasked with developing algorithms for financial sector companies, described how she became aware of the impact of her work: "The figures in my models at the hedge fund stood for something. They were people's retirement funds and mortgages. In retrospect, this seems blindingly obvious. And of course I knew it all along but I hadn't truly appreciated the nature of the nickels, dimes and quarters that we pried loose with our mathematical tools."[20]

Platform leaders are in denial about the extent of the problem. When they come under criticism, they tend to highlight social media's positive contributions. As late as November 2016, Mark Zuckerberg denounced the suggestion that fake news on

Facebook had an impact on the U.S. presidential election as a “crazy idea,” although he later said he regretted the statement.[21]

The algorithms at the heart of social media’s pathologies have received alarmingly little media attention. That needs to change. We need to fundamentally rethink our views on social media platforms.

Transparency: A Strategic Framework for Countering Mobocracy

"Sunlight is said to be the best of disinfectants; electric light the most efficient policeman."
— Former U.S. Supreme Court Justice Louis D. Brandeis[22]

Platforms should no longer use engagement-driven algorithms to maximize revenues at the expense of social well-being. To get accountability, we need far more transparency of the outputs produced by these algorithms. Because algorithms seek to give each user a hyper-personalized experience, presenting individuals with whatever content they are most likely to engage with, we have little idea what other users are consuming. Because we don’t know users’ collective experience and the information that sites are circulating, we can’t hold accountable those who spread fake news and misinformation until it’s too late. The extent to which platforms reward polarization and sensationalism also is invisible to users.

Transparency can deter bad actors from manipulating the system. Moreover, if scholars, journalists, and other interested parties have access to output data, they can help understand the scope and nature of the problem. The data would be a means of holding social media platforms accountable for their impact on society.

The data that social media companies currently share with researchers and other interested parties is insufficient for two reasons. First, platforms have complete discretion about the data they share. Platforms are under no obligation to share data and are not generally interested in collaborations that would result in criticism. During the 2016 U.S. presidential election, Facebook initially refused to release political campaign data citing trade secrets. It was only after mounting public and media pressure that they announced plans to release the data.[23]

Second, the data that platforms do share with researchers is inadequate. Consider, for example, content that is reported by users for breaching a platform’s terms of service and is subsequently removed by the platform. Platforms currently do not disclose information connected to deleted content. It’s impossible to know which information has been deleted, why it was deleted, who posted it, and the time and reach it had before being deleted. Platforms don’t like to share this information because of the potential public embarrassment it could cause.

Towards A Solution: Public Interest APIs

Recent steps by Twitter[24] and Facebook[25] to improve transparency is a move in the right direction, but do not go far enough.

We believe that all platforms using algorithms to distribute content should develop a standardized public interest API (a standard interface for sharing and accessing data) that provides a detailed overview of the information distributed on their networks, while respecting concerns for user privacy, trade secrets, and intellectual property. Social media companies already use aggregate data as a means to alter their own algorithms, introduce new product features, and define the company's strategy.

There are three categories of data that need to be shared:

Public Posts. Malignant actors spread misinformation and manipulate users on the assumption that they can engage large audiences without being held accountable. To counter this, platforms should make data available for all public posts, whether created by an individual user, group, or page. This data need to include reach and engagement figures and provide a demographic breakdown of the audience.

The API should also disclose the top trending stories in different geographic and demographic groupings, and identify the influencers that enabled a public post to achieve viral status. This will allow third parties, like journalists, researchers, and citizens themselves, to identify trends surfacing across the platform, and hold platforms accountable for responding to harmful trends in real time.

Advertising Campaigns. Long gone are the days of uniform print and broadcast media where all advertisements were clearly visible. With the advent of micro-targeting and "dark ads" on social media, we no longer know who is distributing what information to whom. [26] This has huge implications for our ability to detect false advertising, political smear campaigns, and election manipulation. Platforms need to reveal through the public interest API who is purchasing ads, which groups they are targeting, and the content of these ads.

Censored Content. All social media platforms have policies for censoring content that violates their terms of usage. Even so, their algorithms distribute prohibited content for an unknown period of time before it is discovered and removed.

Platforms can use the public interest API to reveal the contents and origins of deleted material that is not targeting specific individuals, the amount of time the content was distributed on the platform prior to deletion, and what reach and engagement the material achieved during that time. Public access to this data will pressure social media companies to move more quickly to delete content that violates their policies and to preserve it for future research. It will also enable journalists and researchers to

understand what sorts of content are most frequently censored and verify that platforms aren't censoring content beyond what their policies claim they censor.

Transparency can no longer be a choice. We need a radical shift in the tech industry's approach to how we communicate with one another. It is no longer acceptable to blindly build products, which carry huge implications for society, without accompanying transparency. We hope that the model for a public interest API presented here will contribute to the conversation on making social media platforms a positive force. Although transparency isn't a full answer, its absence is an obstacle to a sustainable solution. The leaders of Facebook, YouTube, and Twitter have all proclaimed the value of transparency. It is now time for these companies to match their words with commensurate actions. If not, it's only a matter of time before governments intervene with regulatory action.

Areas for Further Research

This paper leaves open a number of questions for further research:

1. What is the impact of engagement-driven algorithms on the behavior of content creators?
2. What would be the design of algorithms that create disincentives to the creation and dissemination of sensational, fake, and polarizing content?
3. How can platforms account for veracity and values in distribution decisions? How can signals for these values be incorporated into discovery algorithms?
4. What are the alternatives to the attention economy's advertising-dependent business model?

Endnotes

[1] Schmidt, Eric, and Jared Cohen. *The New Digital Age: Transforming Nations, Businesses, and Our Lives*. Vintage Books, 2014. pp. 3-4.

[2] Ghosh, Shona, "Sheryl Sandberg Just Dodged a Question About Whether Facebook is a Media Company," Business Insider, October 12, 2017. http://www.businessinsider.com/sheryl-sandberg-dodged-question-on-whether-facebook-is-a-media-company-2017-10

[3] See, for example: Shakya HB, Christakis NA. Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. www.ncbi.nlm.nih.gov/pubmed/28093386

[4] As of 2017, Facebook has grown to over two billion users globally, and financial reports have shown consistent revenue growth since at least 2010 (for more, see https://investor.fb.com/financials/).

[5] Stewart, James B. "Facebook Has 50 Minutes of Your Time Each Day. It Wants More." *The New York Times*, May 5, 2016. www.nytimes.com/2016/05/06/business/facebook-bends-the-rules-of-audience-engagement-to-its-advantage.html

[6] Bruns, Axel et al. *The Routledge Companion to Social Media and Politics*. Routledge, 2016. p. 22.

[7] See, for example, Tufekci, Zeynep, and Christopher Wilson. "Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square." *Journal of Communication*, vol. 62, no. 2, June 2012, pp. 363–379, and Ghonim, Wael. *Revolution 2.0: the Power of the People Is Greater than the People in Power: a Memoir*. Mariner Books/Houghton Mifflin Harcourt, 2013.

[8] Bruns, Axel et al. *The Routledge Companion to Social Media and Politics*. Routledge, 2016. Chapter 1.

[9] Gottfried, Jeffrey, and Elisa Shearer. 2016. "News Use across Social Media Platforms 2016." Pew Research Center, May 26. http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016 Cited in Allcott, Hunt, and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, vol. 31, no. 2, 2017, pp. 211–236., doi:10.1257/jep.31.2.211. The authors note, however, that "only 18 percent report that they get news from social media 'often,' 26 percent do so 'sometimes,' and 18 percent do so 'hardly ever.'"

[10] "Partisan Conflict and Congressional Outreach." Pew Research Center, February 23, 2017. www.people-press.org/2017/02/23/partisan-conflict-and-congressional-outreach/

[11] Shugerman, Emily. “Donald Trump Is Losing Engagement on Twitter, Analysis Shows.” *The Independent*, April 29, 2017. www.independent.co.uk/news/world/americas/us-politics/trump-twitter-less-popular-barack-obama-tweets-north-korea-china-shares-fake-news-a7709871.html

[12] Timberg, Craig. “Russian Propaganda May Have Been Shared Hundreds of Millions of Times, New Research Says.” *The Washington Post*, October 5, 2017. www.washingtonpost.com/news/the-switch/wp/2017/10/05/russian-propaganda-may-have-been-shared-hundreds-of-millions-of-times-new-research-says/?utm_term=.2810a34806f5

[13] Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. “Exposure to Ideologically Diverse News and Opinion on Facebook,” *Science* 348(6239): 1130–32. Cited in Allcott and Gentzkow.

[14] Bishop, Bill. *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*. Houghton Mifflin Harcourt, New York, 2008.

[15] Allcott and Gentzkow.

[16] Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Books, 2012. p. 88.

[17] “Churchill and the Commons Chamber.” UK Parliament, www.parliament.uk/about/living-heritage/building/palace/architecture/palacestructure/churchill/

[18] Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2015.

[19] Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Books, 2012. p. 14.

[20] O'Neil, Cathy. *Weapons Of Math Destruction: How Big Data Increases Inequality And Threatens Democracy*. Crown Publishing, 2017. p. 35.

[21] Levin, Sam. “Mark Zuckerberg: I Regret Ridiculing Fears over Facebook's Effect on Election.” *The Guardian*, September 27, 2017. www.theguardian.com/technology/2017/sep/27/mark-zuckerberg-facebook-2016-election-fake-news

[22] Brandeis, Louis D. *Other People's Money: And How The Bankers Use It* (Classic Reprint). Forgotten Books, 2015.

[23] Ingram, David. “Facebook to Keep Wraps on Political Ads Data despite Researchers' Demands.” Thomson Reuters, June 22, 2017. www.reuters.com/article/us-usa-politics-

facebook/facebook-to-keep-wraps-on-political-ads-data-despite-researchers-demands-idUSKBN19D1CN

[24] Wang, Selina. “Twitter Is Making Its Political Advertising More Transparent.” Bloomberg, October 24, 2017. www.bloomberg.com/news/articles/2017-10-24/twitter-adopts-advertising-transparency-rules-amid-russia-probe

[25] Mark Zuckerberg, Facebook. *Mark Zuckerberg - When Someone Buys Political Ads on TV or…*, www.facebook.com/zuck/posts/10104133053040371?pnref=story

[26] Hern, Alex. “Facebook 'Dark Ads' Can Swing Political Opinions, Research Shows.” *The Guardian*, July 31, 2017. www.theguardian.com/technology/2017/jul/31/facebook-dark-ads-can-swing-opinions-politics-research-shows