WHAT'S THE FUTURE AND WHY IT'S UP TO US

TIM O'REILLY

HARPER BUSINESS
An Imprint of HarperCollins*Publishers*

10

MEDIA IN THE AGE OF ALGORITHMS

AFTER THE 2016 US PRESIDENTIAL ELECTION, THERE WAS A lot of finger-pointing, and many of those fingers pointed at Facebook, arguing that its newsfeed algorithms played a major role in spreading misinformation and magnifying polarization.

False stories claiming that Pope Francis had endorsed Donald Trump, that Mike Pence had said that Michelle Obama was "the most vulgar First Lady we've ever had," and that Hillary Clinton was about to be indicted were shared more than a million times. All were cooked up by Macedonian teens out to make a buck. The story about the "FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide"—also totally fake but shared half a million times—was the work of a Southern California man who started in 2013 to prove how easily disinformation spread, but ended up creating a twenty-five-employee business to churn out the stuff.

Facebook users were not the only ones spreading these stories. Many of them circulated by email and on Twitter, on YouTube, on reddit, and on 4chan. Google surfaced them in Google Suggest, the drop-down recommendations that appear for every user as they begin to type a query.

But it was Facebook that became the focus of the discussion, perhaps because at first Mark Zuckerberg denied the problem, saying in an onstage interview at the Techonomy conference a few days after the election that it was "a pretty crazy idea" that the stories had influenced the outcome. They were a tiny proportion of the total content shared on the site, he argued.

Fake news is the stuff of tabloids. Marginal, once the subject of ridicule. How could it come to play such a large role in shaping our collective future?

At the very least, the 2016 US presidential election showed what Eli

Pariser had called "the filter bubble" in full force. Social media algorithms, driven by "likes," show each person more of what they respond to positively, confirming their biases, reinforcing their beliefs, and encouraging them to associate online with like-minded people. The *Wall Street Journal* created an eye-opening site called Blue Feed/Red Feed that used Facebook's own research data on the political preferences of its users to create side-by-side live feeds of hyperpartisan stories shown to each group. It is shocking just how different the news shown to "extremely liberal" and "extremely conservative" viewers turns out to be. I'd experienced a version of that myself in the stories that were shared with me by my conservative family members, and the progressive stories that I'd shared with them in return. We are living in different worlds. Or perhaps we are just living in a new "post-truth" world, where appeals to emotion carry more weight than facts.

The democratization not just of media distribution but also of its creation played a major role. Colin Megill, founder of pol.is, a service focused on creating better public dialogue, told me that his mother, a doctor who worked her whole life to break the glass ceiling, was beset by doubt about Hillary Clinton and had been especially influenced by a video claiming that her aide Huma Abedin had been a member of the Muslim Brotherhood, a video that had autoplayed after she watched YouTube replays of late-night television.

"I reflected on my conversation with my mom a lot after that happened and came up with one possible explanation," Colin said. "For her whole life, something would be out of the news immediately if it was totally false. Editors saw to that. The idea that something with a high production value, shared by millions, could be without a shred of truth really wasn't in her matrix of possibilities." The notion that the video could have been created by an anonymous Trump supporter was just not part of her mental map.

According to Pew Research, 66% of Americans get their news through social media sites, 44% of them from Facebook alone. Much of that content may come from traditional media via links shared on social media, but much of it is native to the platform, or coming from new, hyperpartisan sites like those cooked up for profit by the Macedonian

teens, or for partisan reasons by extreme right-wing or extreme left-wing political organizations. And that is to say nothing of groups like ISIS that have successfully used social media for terrorist recruiting, or of the role of propaganda planted or amplified by Russia with the goal of influencing the US presidential election. As one US government official who wished to remain anonymous told me: "We aren't fighting our first cyberwar. We just fought it. And we already lost."

ALGORITHMIC WHAC-A-MOLE

In many ways, the rising influence of fake news is a cautionary tale of algorithms gone wrong, digital djinns given poorly framed instructions with potentially catastrophic consequences. It is worth studying even though Facebook and Google will have done a great deal of work to solve the current iteration of the problem by the time this book is published.

In a follow-up Facebook post the week after his dismissive comments, Mark Zuckerberg admitted that fake news was a problem, and that Facebook was working on it. His suggested solution was to give "the community" more tools for signaling what they believed to be true or false. I had met with Mark a few weeks before the election, about a related issue he was wrestling with, how Facebook could give voice to its users around community norms and values. His goal to make Facebook a neutral platform through which its users can connect and share is deeply felt. In his post about fake news and the election, he concluded, "In my experience, people are good, and even if you may not feel that way today, believing in people leads to better results over the long term."

That belief that controlling fake news should be up to the users, not to the platform, shaped Facebook's response to the crisis. Mark wrote: "We have already launched work enabling our community to flag hoaxes and fake news, and there is more we can do here. We have made progress, and we will continue to work on this to improve further." So far, so good.

He continued to argue for the role of Facebook's users in policing

the site: "I am confident we can find ways for our community to tell us what content is most meaningful, but I believe we must be extremely cautious about becoming arbiters of truth ourselves." He correctly noted that "identifying the 'truth' is complicated. While some hoaxes can be completely debunked, a greater amount of content, including from mainstream sources, often gets the basic idea right but some details wrong or omitted. An even greater volume of stories express an opinion that many will disagree with and flag as incorrect even when factual."

The internal debate at platforms such as Facebook and Google about their responsibility to control fake news is not just a matter of caution in getting it right, though. It's also a worry about setting a legal precedent. The Digital Millennium Copyright Act (DMCA), enacted in 1998, exempted Internet service providers and other online intermediaries from liability from copyright infringement on the grounds that they were neutral platforms that simply enabled users to post whatever they want. They are more like a wall on which users can post handbills than they are like a publisher who chooses what to publish and should be held to a higher legal standard. This "neutral platform" argument is central to the existence of Internet services. Without it, Google would be liable for every copyright infringement made by any user posting online, simply by including that content in the search index. Similarly, Facebook, Twitter, YouTube, or WordPress would be liable if any user posted infringing material. A similar legal defense, by extension, could be applied to other kinds of content posted by users: The service is a platform for its users, not a content provider. No online service wants to break this shield.

Critics snarl at this defense. One such critic, Carole Cadwalladr, was outraged that Google's Suggest feature was offering results such as "Jews are evil" as autocomplete for "Jews are . . ." When she clicked through, she found that the first result had the headline: "Top 10 Major Reasons Why People Hate Jews." A page from neo-Nazi site Stormfront was the third result, with additional explanations of why Jews are evil appearing as the fifth, sixth, seventh, and tenth results. When she did a search for "did the holo . . ." Google autocompleted her query to

"did the Holocaust happen?" and she was taken to a list of Holocaust-denial sites, again topped by a page from Stormfront.

Her solution: Google should stop linking to these pages immediately. "Google's business model is built around the idea that it's a neutral platform. That its magic algorithm waves its magic wand and delivers magic results without the sullying intervention of any human," she wrote in a scathing op-ed for the *Guardian*. "It desperately does not want to be seen as a media company, as a content provider, as a news and information medium that should be governed by the same rules that apply to other media. But this is exactly what it is."

I sympathize with Cadwalladr's outrage, and her belief that Google (like all media) "frames, shapes and distorts how we see the world." I agree that Google needs to come to grips with bad results like this, just as they have come to grips with other challenges to the quality of their results. But Cadwalladr ignored the scale at which Google operates, and the way that scale fundamentally changes the necessary solution.

> **Google, Facebook, Twitter, and their like need to be understood as a new thing, which doesn't fit neatly into the old map. That new thing operates by different rules—not by whim or an unwillingness to incur the costs of curation, but by necessity.**

Google's and Facebook's reluctance to make manual interventions is not just a matter of hiding behind a convenient legal disclaimer of responsibility. These sites don't produce their results through some convocation of human editors, like the old *New York Times* front-page meeting, in which editors decided which stories get placement and where. That meeting was phased out even at the *Times* in 2015. The result of any Google search is the result of prodigious efforts to retrieve and rank every page on the web—30 *trillion* of them, from 250 billion unique web domain names, according to former Google VP of search Amit Singhal—and to serve them up in response to more than 5 billion searches a day. Many of those searches are common, but at least tens of millions of them are the result of quite infrequent combinations of words and phrases. The offensive Holocaust results that Cadwalladr

complained about are the result of a search that, according to Google, is made only about 300 times a day. Out of 5 billion. That's 0.000006% of daily searches, a few millionths of a percent.

Facebook is similarly huge. In 2013, the social network disclosed that nearly 5 billion pieces of content were posted every day. That number is now surely far larger, as the site now has over 1 billion daily active users, up from 700 million in 2013.

The idea that Google or Facebook can solve the problem simply by hiring teams of human editors or fact checkers, or use outside media organizations to combat fake news, hate speech, or other objectionable results, removing or demoting them one at a time, indicates that people have little idea of the scale or nature of the problem. It's like the carnival game of Whac-A-Mole, except with billions of moles and only hundreds of hammers. Human oversight and intervention is definitely needed, but it will make little difference if it is implemented in the way that critics like Cadwalladr imagine. To whack billions of moles, you need much faster hammers.

We have to break the notion that the role of the human in the loop is as the final decision maker pulling a kill switch. There's a famous *Harvard Business Review* article called "Who's Got the Monkey?" that explains why whenever an employee brings in a problem, like a monkey on his or her back, the manager must offer counsel, and then send the employee back out with the monkey. Otherwise, the manager, with multiple employees, ends up with all the monkeys. How much more true is this in the age of algorithms? The manager ends up with a million monkeys. A good manager is always a teacher. How much more is this true with the powerful but fundamentally stupid race of djinns that do so much of the work at our massive online platforms?

Google no doubt has teams of developers, the managers of the digital workers who build the index and serve up the search results, hard at work teaching their inhumanly fast djinns how to mitigate this problem. I'd be very surprised if, by the time this book has been published, there hasn't been a comprehensive fake news search overhaul akin to the 2011 Panda and Penguin updates that dealt with content farms. And in fact, within weeks of Cadwalladr's op-eds, the search results for

Holocaust denial had been improved. The initial fix had failed to work consistently, and Google is still struggling to come up with a comprehensive solution to fake news, but the processes by which they respond to attacks on the search engine's effectiveness are well defined.

Facebook's problems are not identical to Google's. While Google evaluates and links to content from hundreds of billions of external sites, Facebook's content is posted natively by its users on its own platform. Much of that content links to external sites, but much of it does not. Even when the content comes from external sites, it has often been remixed into a "meme"—which has now come to mean a graphic or video representation of a key moment or quote that is freed from its original context, designed to be shared, designed for impact rather than deeper dialogue or understanding.

In May 2016, long before Trump was elected, Milo Yiannopoulos, writing on *Breitbart*, predicted that Trump's facility with creating Internet memes and appealing to the people who share them was crucial to his success. "Establishment types no doubt think this is all silly, schoolyard stuff," he wrote. "And it is. But it's also effective. . . . Caught between the hammer of Trump's media machine and the anvil of his online troll army, The Donald's opponents never stood a chance. Trump understands the Internet, and the Internet might just propel him into the White House. Meme magic is real."

As a result of the lack of context, many of the signals that Google relies on, such as the link structure of the web, are absent. While Facebook can make use of some of the same techniques, its infrastructure and business processes for dealing with content are not the same. This is one reason that Facebook is looking for "the community" to solve the problem. Can its billion-plus users police the site given the right tools? In a patent filed in June 2015, System and Methods for Identifying Objectionable Content, Facebook had already laid out its approach to dealing with hate speech, pornography, and bullying, relying on user reporting but using many additional signals to rank and weight not only the reports themselves but the users providing them. Many of the techniques described in the patent are also applicable to fake news.

In a second blog post on the topic, Mark Zuckerberg wrote in more detail about the company's approach, which includes making it easier for people to report fake stories, partnering with third-party fact-checking organizations, and potentially even showing warnings on stories that have been flagged by fact checkers or the community. But Mark also pointed out that the most important thing Facebook can do is "to improve our ability to classify misinformation. This means better technical systems to detect what people will flag as false before they do it themselves." He also noted that Facebook had already improved the algorithms used to choose "related articles" under links in the News Feed.

This algorithmic reeducation is essential because the speed with which content can spread on social media works against unaugmented human fact checkers. One fake story began on Twitter when Trump supporter Eric Tucker posted a photo of chartered buses in Austin, Texas, and suggested that the Clinton campaign was using them to bus protesters to Trump's upcoming speech. Even though Tucker himself had only forty followers, and deleted the tweet once he found that the buses were actually for visitors to a convention held by software company Tableau, the photo went viral, shared 16,000 times on Twitter and 350,000 times on Facebook. His initial tweet had used the hashtags #fakeprotests #Trump2016 #Austin, ensuring that it would be read widely by people following those topics.

The story was picked up first on reddit, then by various right-wing blogs, and then by mainstream media. Donald Trump himself then tweeted about "professional protesters," adding fuel to the fire. While Tucker didn't expect to have such an impact, the people who promote fake news often have strong incentives to boost it, using programmatic tools to discover key influencers and plant it with them to give it a quick start. Given the traffic that a hot story can bring today, even professional news organizations use automated "social listening tools" to quickly pick up trending topics and republish popular stories on their own publications without the careful fact checking that used to characterize mainstream media.

By the time concerned users or fact checkers begin to flag content as

false, it may already have been shared hundreds of thousands of times and have been read by millions. Retractions of the original story usually have little effect. By midnight of the day he first tweeted it, Tucker had deleted the original tweet and replaced it with one stamped "False" across the picture. That tweet was retweeted a grand total of 29 times, versus the 16,000 retweets of the original. I'm reminded of the old saying passed on to me by my mother: "A lie will have gone halfway around the world before the truth has had time to tie on its shoes."

One approach that Google, Facebook, and others have begun practicing, labeling disputed stories, may help, because the labels will follow and potentially stay with the story, but only if it's done in advance of the story being too widely shared. But even this approach has problems, since there is nothing to stop a partisan or financially motivated site from creating a new version of the same false story. How do you detect that? You're back to the algorithmic djinns for help whacking the mole.

In addition, users themselves have trouble not only determining what is true or false, but even in detecting the signals that companies provide to help them determine the authority of what they are seeing. Only 25% of high school students in one Stanford study recognized the significance of the blue check mark used by Facebook and Twitter to denote verified accounts. Will flags for fake news fare any better?

Finally, it's essential to realize that search engines and social media platforms are the battlefield of an online war, with hostile attackers using the same tools that were originally developed by advertisers to track their customers, and then by scammers and spammers to game the system for profit. In addition to the Russian-sponsored social media disinformation campaigns, the Trump campaign's Project Alamo used highly targeted disinformation to discourage Clinton voters from going to the polls. These posts were referred to as "dark posts" by Brad Parscale, who led the campaign's social media efforts, private posts whose viewership is tightly targeted so that, as he put it, "only the people we want to see it, see it."

Jonathan Albright, a communications professor who analyzed a network of 300 news sites that were promulgating fake news during the 2016 election, made the same point about programmatic microtargeting.

"This is a propaganda machine," he wrote. "They're capturing people and then keeping them on an emotional leash and never letting them go."

"Capturing people and then keeping them on an emotional leash" is nothing new. It was at the heart of much media in the days of "yellow journalism" at the turn of the twentieth century, beaten back by journalistic standards for much of the century, then reasserted in its closing decades by talk radio and by Fox News on TV. Social media and its advertising business model has taken the process to its logical conclusion.

Targeted social media campaigns will almost certainly be a feature of all future political campaigns. Online social media platforms—and society as a whole—will need to come to grips with the challenges of the new medium. The moment of crisis may come when we realize that the tools of disinformation and propaganda are the very same tools that are routinely used by businesses and ad agencies to track and influence their customers. It is not just political actors who have a vested interest in spreading fake news. Vast sums of money are at stake, and participants use every tool to game the system. The problem is not Facebook's.

> **Fake news is simply the most unsavory face of the business model that drives much of the Internet economy.**

In cybercrime, these tools go beyond the distasteful into the realm of the illegal. One Russian botnet uncovered in December 2016 was creating targeted videos that were generating $3–5 million per day in ad revenue from fake video views by programs masquerading as users. In other words, this battle goes far beyond planting fake news. It is also possible to plant fake users who exist only as imaginary pawns in a battle of clicks and likes.

When attackers use programs to masquerade as users, unaided human supervision is inadequate due to the speed and scale of the attacks. This is another reason why the response to fake news and other kinds of amplified social media fraud needs to be algorithmic, much as spam filters are, rather than solely relying on users or the tools of traditional journalism.

The 2015–16 DARPA Cyber Grand Challenge was based on a similar insight, asking for the development of AI systems to find and automatically patch software vulnerabilities that corporate IT teams just aren't able to keep up with. The problem is that an increasing number of cyberattacks are being automated, and these digital adversaries are finding the holes far faster than humans can patch them.

John Launchbury, the director of DARPA's Information Innovation Office, told me an illuminating story from the Cyber Grand Challenge. The various competing systems had been seeded with security vulnerabilities that they were expected to find and fix before they could be exploited by another of the systems. One of the AI contestants examined its own source code and found a vulnerability not among those that had been planted, and used it to take control of another system. A third system, observing the attack, diagnosed the problem and fixed its own source code. All of this in twenty minutes.

Air Force Colonel John Boyd, "the father of the F-16," introduced the term *OODA loop* ("Observe-Orient-Decide-Act") to describe why agility is more important in combat than pure firepower. Both fighters are trying to understand the situation, decide what to do, and then act. If you can think more quickly, you can "get inside the OODA loop of your enemy" and disrupt his decision making.

"The key is to obscure your intentions and make them unpredictable to your opponent while you simultaneously clarify his intentions," wrote Boyd's colleague Harry Hillaker in his eulogy to Boyd. "That is, operate at a faster tempo to generate rapidly changing conditions that inhibit your opponent from adapting or reacting to those changes and that suppress or destroy his awareness. Thus, a hodgepodge of confusion and disorder occur to cause him to over- or under-react to conditions or activities that appear to be uncertain, ambiguous, or incomprehensible."

This is very hard to do when your opponent is a machine able to act millions of times faster than you are. One observer who wished to remain anonymous, an expert in both financial systems and in cyberwarfare, said to me, "It takes a machine to get inside the OODA loop of another machine."

WHAT IS TRUTH?

We have been talking about objectively verified facts and objectively verified falsehoods. There is a further, even more challenging problem that algorithms can be unexpectedly helpful with. As Mark Zuckerberg noted, many problematic pieces of content are not outright falsehoods, but contain opinion or half-truths. Partisans on both sides of an issue are eager to believe and reshare content even if they know it is at least partially false. Even when professional fact-checking organizations such as Snopes or PolitiFact or mainstream media sites staffed by experienced reporters debunk a story, there are others who decry the result as biased.

George Soros has pointed out that there are things that are true, things that are false, and things that are true or false only to the extent that people believe in them. He calls this "reflexive knowledge," but perhaps the old-fashioned term *beliefs* will serve just as well. So much that matters falls into this category—notably history, politics, and markets. "We are part of the world we seek to understand," Soros wrote, "and our imperfect understanding plays an important role in shaping the events in which we participate."

This has always been the case, but our new, world-spanning digital systems, connecting us into a nascent global brain, have accelerated and intensified the process. It is not just facts that spread from mind to mind. It is not just the idea that pots containing decaffeinated coffee should be orange. Misinformation goes viral too, shaping the beliefs of millions. Increasingly, what we know and what we are exposed to are shaped by personalization algorithms, which try to pick out for us from the firehose of content on the Internet just the things that the algorithms expect we will most likely respond to, appealing to engagement and emotion rather than to literal truth.

But Soros's reminder that stock prices and social movements are neither true nor false suggests an approach to the fake news problem as well. Even while recognizing the role of emotion in stock prices, stock pickers still believe that a stock has "fundamentals." A stock price may depend on what people believe about a company's future prospects, but

they recognize that a company also does have revenue, income, capital, growth rates, and a plausible market opportunity from which those future prospects can be estimated. Stock reporting routinely measures and reports on the price/earnings ratio and other measures of how far expectations outstrip the fundamentals, so that people can make informed judgments of how much risk they are taking. There are many who will overlook the risks, and those who encourage them to do so, but at least some information is there.

> **The distance between human enthusiasm and the fundamentals can also be measured for news, using many signals that can be verified algorithmically by a computer, often more quickly and thoroughly than they can be verified by humans.**

When people are discussing the truth or falsity of news, and the responsibility of sites like Facebook, Google, and Twitter to help identify it, they somehow think that determining "truth" or "falsity" is solely a matter of evaluating the content itself, and make the case that it can't be done by a computer because it requires a subjective judgment. But as with Google Search, many of the signals that can be used are independent of the actual content. To use them, we must simply follow Korzybski's injunction to compare the map with the territory it claims to describe.

> **Algorithmic fact checking doesn't replace human judgment. It amplifies our power to exercise it, in much the same way as earthmoving equipment amplifies our muscles. The signals it uses are similar to those that a human fact checker might use.**

Does the story or graph cite any sources? If no sources are given, it is far from certain that the story is false, but the likelihood increases that it should be investigated further. A fake story typically provides no sources. For example, when debunking one claim sent to me by my brother, a fake map purporting to show higher crime rates in precincts that voted Democratic, I was unable to find any sources for the data

the map claimed to be based on. In the course of my search, though, I found a series of visualizations put together by *Business Insider* that painted a very different picture. Unlike my brother's map, the legitimate publication provided the source of the data it had used, an FBI crime database.

Do the sources actually say what the article claims they say? It would have been entirely possible for *Business Insider* to claim that the data used in their article was from the FBI, but for there to be no such data, or for the data there to be different. Few people trace the chain of sources to their origin, as I did. Many propaganda and fake news sites rely on that failure to spread falsity. Checking sources all the way back to their origin is something that computers are much better at doing than humans.

Are the sources authoritative? In evaluating search quality over the years, Google has used many techniques. How long has the site been around? How often is it referenced by other sites that have repeatedly been determined to be reputable? Most people would find the FBI to be an authoritative source for US national crime data.

If the story references quantitative data, does it do so in a way that is mathematically sound? For example, anyone who has even a little knowledge of statistics will recognize that showing absolute numbers of crimes without reference to population density is fundamentally meaningless. Yes, there are more crimes committed by millions of people in New York City or Chicago than by hundreds in an area of rural Montana. That is why the FBI data referenced by the *Business Insider* article, which normalized the data to show crimes per 100,000 people, was inherently more plausible to me than the fake electoral maps that set me off on this particular quest for truth. Again, math is something computers do quite well.

Do the sources, if any, substantiate the account? If there is a mismatch between the story and its sources, that may be a signal of falsity. Even before the election, Facebook had rolled out an update to combat what they call "clickbait" headlines. Facebook studied thousands of posts to determine the kind of language typically used in headlines that tease the user with a promise that is not met by the content of the actual

article, then developed an algorithm to identify and downgrade stories that showed that mismatch. Matching articles with their sources is a very similar problem.

Are there multiple independent accounts of the same story? This is a technique that was long used by human reporters in the days when the search for truth was properly central to the news. A story, however juicy, would never be reported on the evidence of a single source. Searching for multiple confirming sources is something that computers can do very well. Not only can they find multiple accounts, but they can also determine which ones appeared first, which ones represent duplicate content, how long the site or username from which the account has been posted has existed, how often it makes similar posts, and even which location the content was posted from.

Consumers of online media are unlikely to retrain themselves to act this same way. Especially when they read a story that confirms their biases, few people do a search for other accounts of the same story from a source that doesn't share those biases. One of my sisters sent me a story about California "legalizing child prostitution" after reading an account in the *Washington Examiner*. "I think this might just be why some decent people don't like California," she wrote. I read the bill, as well as rebuttals from other media sources. What the California bill actually said was that individuals under the age of eighteen involved in prostitution would not be treated as criminals, but instead could be taken into custody and made a ward of the court. Given an account of an original source, an algorithm could potentially compare the summary with the original, or compare multiple accounts of the same event, and flag discrepancies.

In addition to sharing content that confirms their biases and framing it to serve their agendas, users are too eager for clicks and likes. John Borthwick, CEO of Betaworks, described the user behavior that feeds the spread of false news. "Media hacks take advantage of the decontextualized structure of real-time news feeds," he wrote. "You see a Tweet from a known news site, with a provocative headline and maybe the infographic image included—you retweet it. Maybe you intend to read the story, might be you just want to Tweet something interesting

and proactive, maybe you recognize the source, maybe you don't." One of the simplest algorithmic interventions Facebook and Twitter could make would be to ask people, "Are you sure you want to share that link? You don't appear to have read the story."

Because they follow rules exactly, algorithms are also good at noticing things that slip by humans. Earlier in this chapter, I cited an op-ed by Carol Cadwalladr about Google and Holocaust denial sites. At the end of a follow-up article, in which Cadwalladr showed how she could push down the fake results by buying a few targeted ads, was an explanation attributed to Danny Sullivan, the search engine guru, saying that Google had changed its algorithms "to reward popular results over authoritative ones. For the reason that it makes Google more money."

The article seemed doubly authoritative—it appeared in the *Guardian*, a reputable newspaper, and it quoted an expert on Google search I know and respect. But something was nagging at me. While there were other links in the op-ed, there was no link to the article from which Danny Sullivan was supposedly quoted. So I sent Danny an email. He told me that not only had he not said that Google had changed its algorithm to increase its profits, but he'd notified the *Guardian* after the article cited him incorrectly. Sadly, he said, the article hadn't been updated.

> **Citing and linking to sources makes it much easier to validate whether an assertion is an opinion or interpretation, and who is making it. This should be the gold standard for all reporting. If media reliably linked to sources, any story without sources would automatically become suspect.**

There are cases, of course, where reporters depend on anonymous sources. Watergate's "Deep Throat" comes to mind. But note how journalistic standards have slipped: Woodward and Bernstein spent many months tracking down corroborating evidence that proved Deep Throat's assertions. They didn't just report the leaked information as hearsay.

REASONABLE DOUBT

When fake news is detected, there are a number of possible ways to respond.

The stories can be suppressed entirely if certainty is extremely high. This should be done rarely, because suppressing content entirely is a slippery slope toward censorship. We already rely on this level of extreme prejudice in other online applications, though, since it is what email providers do to filter the email we actually want to see from the billions of spam messages sent every day.

The stories can be flagged. For example, Facebook (or online mail systems like Gmail, since much fake news appears to be spread by email) could show an alert, similar to a security alert, that says, "This story appears likely to be false. Are you sure you want to share it?" with a link to the reasons why it is suspect, or to a story that debunks it, if that is available. Unfortunately, Facebook's desire not to be the arbiter of truth, even when the stories are from known sources of misinformation, means that their efforts are often less effective than they could be.

In March 2017, Facebook began listing stories as "disputed" when authorized sites like Snopes or PolitiFact debunk them, but as expected with human fact checkers, the process takes days when the damage is done in minutes or hours. Krishna Bharat, the Google engineer who founded and ran Google News for many years, believes that one of the most important roles for algorithms to play may be as a kind of circuit breaker, which pauses the spread of suspicious postings, providing "enough of a window to gather evidence and have it considered by humans who may choose to arrest the wave before it turns into a tsunami." Bharat points out that it is not every false story that needs to be flagged, only those that are gaining momentum. "Let us say that a social media platform has decided that it wants to fully address fake news by the time it gets 10,000 shares," he notes. "To achieve this they may want to have the wave flagged at 1,000 shares, so that human evaluators have time to study it and respond. For search, you would count queries and clicks rather than shares and the thresholds could be higher, but the overall logic is the same."

A variation of Facebook's existing automated Related Stories feature

might be another way to tackle confirmation bias without resorting to blocking a story entirely. Given a news story that displays likely bias according to various algorithmic measures, it should be possible to match it up immediately with an offsetting story from a site known to be authoritative, or to match it up with original sources. While nothing will force readers to consult those sources, the fact that a story is flagged as potentially false or misleading and that an alternative view is available may give pause to the trigger finger of sharing. But this has to happen extremely quickly, before content has already gone viral.

Suspect stories also can be given less priority, shown lower down in the newsfeed, or less often. Google does this routinely in ranking search results. And while the idea that Facebook should do this has been more controversial, Facebook is already ranking stories, featuring those that drive more engagement over those that are more recent, showing stories related to ones we've already shared or liked, and even showing particularly popular stories more than once. Once Facebook stopped showing stories in pure timeline order, they put themselves in the position of curating the feed algorithmically. It's about time they added source verification and other "truth" signals to the algorithm.

The algorithm does not have to find absolute truth; it has to find a reasonable doubt, just like a human jury. This is especially true if the penalty is simply not being promoted. There is no free speech obligation for platforms to proactively promote any particular content. Fake news got a big boost from a flawed algorithm that seems to have favored the emotional rush of partisan engagement over other factors.

Google and Facebook constantly devise and test new algorithms. Yes, there is human judgment involved. But it is judgment applied to the design of a system, not to each specific result. Designing an effective algorithm for search or the newsfeed has more in common with designing an airplane so it flies than with deciding where that airplane flies.

In the case of making an airplane fly, the goals are simple—stay aloft, go faster, use less fuel—and design changes can be rigorously tested against the desired outcome. There are many analogous problems in search—finding the best price, or the most authoritative source of information on a topic, or a particular document—and many that are

far less rigorous. When users get right to what they want, the users are happy, and so, generally, are advertisers. Unfortunately, unlike search, where the desires of the users to find an answer and get on with their lives are generally aligned with "give them the best results," prioritization of "engagement" may have led Facebook in the wrong direction. Engagement and time on-site may be good for advertisers; they may not be good for users or for seekers of truth.

Even in the case of physical systems like aerodynamics and flight engineering, there are often hidden assumptions to be tested and corrected. In one famous example that determined the future of the aerospace industry, a radically new understanding of how to deal with metal fatigue was needed. At the beginning of commercial jet travel, in 1953, Britain's new de Havilland Comet was ready to dominate the skies. Then, horrifyingly, one of the planes fell out of the sky for no apparent reason. The airline blamed pilot error and bad weather. A year later, the skies were clear when a second plane did the same thing. The fleet was grounded for two months during an extensive investigation, after which the manufacturer confidently asserted that they had made modifications to deal with "every possibility that imagination has suggested as a likely cause of the disaster." When a third plane fell from the sky only a few days after the report was issued, it was clear that de Havilland's imagination was insufficient to the challenge. A young engineer in America had a better idea, which handed the future of commercial jet aviation to Boeing. As described by University of Texas physics professor Michael P. Marder, who brought this story to my attention: "Cracks were the centerpiece of the investigation. They could not be eliminated. They were everywhere, permeating the structure, too small to be seen. The structure could not be made perfect, it was inherently flawed, and the goal of engineering design was not to certify the airframe free of cracks but to make it tolerate them."

> **So too, the essence of algorithm design is not to eliminate all error, but to make results robust in the face of error. The fundamental question to ask is not whether Facebook should be curating the newsfeed, but how.**

Where de Havilland tried in vain to engineer a plane where the materials were strong enough to resist all cracks and fatigue, Boeing realized that the right approach was to engineer a design that allowed cracks, but kept them from propagating so far that they led to catastrophic failure. That is also Facebook's challenge. Their goal is to find a way for the plane to fly faster, but fly safely. This means improving their algorithms—training and managing their electronic workers rather than throwing them out and simply going back to human curation. After the de Havilland Comet incidents, the airline industry didn't simply throw up its hands, go back to propeller planes, and give up on commercial jet flight. Facebook's algorithms have been set to optimize for engagement; they need to be more complex, and add optimizations for truth.

The bright side: Searching through the possibility space for the intersection of truth *and* engagement could lead Facebook to some remarkable discoveries. Pushing for what is hard makes you better.

There are signs of this effort in Mark Zuckerberg's February 2017 manifesto, "Building Global Community." In it, he pointed to a radically different way of solving the problem. Mark gave only a token nod to the explicit problem of fake news, noting that new AI tools are already submitting a third of all stories sent to Facebook's internal content review team. (The other two-thirds are submitted by Facebook users.) He focused instead on the root cause of the problem: the decline in social capital, the ties that bind us together as a society and that make it easier for us to work together for the common good.

In his 2000 book, *Bowling Alone*, Robert Putnam used the decline of bowling leagues and the rise of individual bowling as a metaphor for the changing nature of American society. From the days when Alexis de Tocqueville first analyzed the American character in the early nineteenth century, the United States had been characterized by a rich civic fabric of participation in local government, churches, unions, mutual aid societies, charities, sports leagues, and associations of all kinds. The decline of this participation had serious consequences, Putnam thought.

During earlier research on economic differences between the twenty

regional governments of Italy, Putnam had noticed that there was a close correlation between civic engagement and prosperity. "These communities did not become civic simply because they were rich. The historical record strongly suggests precisely the opposite: They have become rich because they were civic." Social capital is as important as financial capital in the wealth of nations.

Mark Zuckerberg came to much the same conclusion. "There has been a striking decline in the important social infrastructure of local communities over the past few decades," he noted. "The decline raises deeper questions alongside surveys showing large percentages of our population lack a sense of hope for the future. It is possible many of our challenges are at least as much social as they are economic—related to a lack of community and connection to something greater than ourselves."

Online communities represent a bright spot, Mark noted, but there is much work to do to expand their impact and their scale, using them to enable offline as well as online connection, empowering community leaders with new tools, and identifying more "meaningful groups" that can have a positive effect on people's offline as well as online lives. Support groups for new parents or for those suffering from a serious disease are good examples. (Margaret Levi, the director of the Stanford Center for Advanced Study in the Behavioral Sciences, pointed out to me one major caveat: that these groups already have a pressing common purpose; finding each other is the problem, which Facebook can clearly help with. In other areas, finding a common purpose that brings people together rather than driving them apart is precisely the unsolved problem.)

When Mark says it is time for Facebook to shift from a focus on friends and family to "the social infrastructure for community—for supporting us, for keeping us safe, for informing us, for civic engagement, and for inclusion of all," you can see the promise of a virtuous circle of engagement. Where engagement seems to be the wrong fitness function for traditional ad-supported media, engagement is exactly the metric we want going up and to the right if we are looking to strengthen not only friendship and families but society as a whole.

That is a very promising direction. If Facebook is indeed able to make progress in strengthening forms of positive engagement that actually create communities with true social capital, and is able to find an advertising model that supports that goal rather than distorts it, that would likely have a greater impact than any direct attempt to manage fake news. When tuning algorithms, as in ordinary life, it is always better to tackle root causes than symptoms. Humans are a fundamentally social species; the tribalism of today's toxic online culture may be a sign that it is time to reinvent all of our social institutions for the online era.

In our conversation on the topic, Margaret Levi offered a concluding warning: "Even when social media helps people engage in collective action—as it did in Egypt—by coordinating them, that is quite distinct from an ongoing organization and movement." This is what our mutual friend, Wael Ghonim, had learned as a result of his experience with the Egyptian revolution. "Unanswered still," Margaret continued, "is Wael's concern about how you transform coordinated and directed action to a sustained movement and community willing to work together to solve hard problems. Especially when they begin as a heterogeneous set of people with somewhat conflicting end goals. They may agree on getting rid of the dictator, but then what?"

THE PROBLEM OF DISAGREEMENT

Henry Farrell, a professor of political science at George Washington University and a columnist for the *Washington Post*, wrote to me after reading an online post that I'd published about the fake news problem. Henry made an important point very different from my own. The problem, he wrote, is "[n]ot what is the optimal solution to finding truth given the technology and the constraints. Instead . . . what is the most plausible path towards identifying a sustainable *political* compromise between a heterogeneous crowd of individuals who don't agree on the solution, and in some cases maybe don't agree that there is a problem in the first place?"

This is a very good question, but, I would argue, also one that tech-

nology may be able to help with. In a very interesting experiment, the government of Taiwan held a public consultation, Virtual Taiwan, using a tool called pol.is to involve its citizens in discussions of legislation and regulations, including, notably, regulation of new transportation services such as Uber.

As Colin Megill, the creator of pol.is, describes it, Jaclyn Tsai, a minister in the executive branch in Taiwan, went to a government-oriented hackathon and said, "We need a platform to allow the entire society to engage in rational discussion."

Pol.is asks people to make assertions that take the form of a single-sentence comment. Those reading those assertions don't have a means to argue with them—there are no replies. They can agree, disagree, or pass. And then they can make a separate assertion of their own. Colin notes, "Doing away with replies gets you something very special. It gets you a matrix [of] every participant, and what they thought about every comment." Humans aren't very good at analyzing this, but machines are really good at it. "You use this all the time," he says. "Every time you rate a movie, every time you buy a product, you're creating data; and we do machine learning on that data in pol.is like Netflix would do on movies. Netflix identifies clusters, things like people who love comedy, people who love horror, people who love comedy and documentaries but hate horror, people who love comedies and horror but hate documentaries."

In pol.is, a well-known statistical technique called principal component analysis (PCA) is used to cluster the assertions and the people who respond to them into groups of like-minded individuals and the statements they favor and disfavor. The statements each group tended to vote uniquely on, as well as statements that enjoyed consensus among all the groups, are shown to everyone. The assertions getting consensus across all groups, or within specific groups, float to the top and are seen more often—just like content on Facebook, but with visibility into what percentage of others agreed or disagreed with them.

This is very different from Facebook likes because participants can see the filter bubble–like graph of those who agree and disagree with a common set of assertions. Participants can click through to view the

statements that shape a particular cluster. And as participants agree or disagree with various statements, their avatars move on the graph, toward or away from another cluster. Participants can see not only what percentage of the entire conversation agrees with them on a particular statement, but also the percentage of the cluster who agrees with similar statements they or others have made.

There is a similar, very powerful technique for small groups meeting in the physical world, which we've often used to discuss contentious issues among the staff and fellows at Code for America. It's called a "Human Spectrogram." The group stands together in the middle of a large room. Someone makes a statement, and those who agree strongly with it move to the far end of the room. Those who disagree move to the other end of the room. People whose views are less polarized can arrange themselves anywhere in between. Then someone makes another comment, and if it influences your thinking, you move accordingly. The beauty of pol.is is that it seems to have scaled this approach to work with thousands of people and thousands of assertions across multiple dimensions.

The pol.is discussion of Uber in Virtual Taiwan began with one assertion: "I think Passenger Liability insurance should be mandatory for riders on uberX private vehicles." Those responding to this assertion quickly sorted themselves into groups: those pro- and anti-regulation. Participants could see the size of those groups—no more than 33% took either side of the debate. So people tried out different assertions, trying to move toward those that would garner higher support.

Over a period of four weeks, the group of about 1,700 participants in the Uber conversation (out of tens of thousands who participated in the overall Virtual Taiwan effort) worked their way toward consensus on key points. One assertion that reached high agreement: "The government should leverage this opportunity to challenge the taxi industry to improve their management & quality control system, so that drivers & riders would enjoy the same quality service as Uber. (95%, across all groups.)"

By the end of the consultation, Uber had agreed to provide Minister Tsai with its international liability insurance policy and, if needed, release it for public review. It also agreed to coach all drivers to register

and obtain professional driver's licenses, and that if it were legalized in some areas, it was willing to pay for UberX car permits and transport taxes. The Taipei Taxi Association expressed a willingness to work with the UberTAXI platform, and to offer better services if the government would let them increase taxi pricing in response to market demand in the same way that Uber does.

Ray Dalio, the founder and executive chairman of Bridgewater Associates, uses a similar approach to creating what he calls an "idea meritocracy" at his company, the largest hedge fund in the world. As members of the firm debate investments or ideas, they rate the assertions of the other participants, assembling them into a matrix that highlights agreement and disagreement. Everyone is urged to be "radically transparent" with their opinions, and the newest associate is welcome to tell Ray himself that he is wrong. Bridgewater takes the further step of applying an algorithm to the matrix, which takes into account factors such as past performance, expertise on the particular topic, and other ways of weighting individual opinions. The goal is to combine the best of human insight and the ability of computer algorithms to sum up and clarify the points of agreement and disagreement.

There's no silver bullet, and disagreement too can be a tool for moving toward truth, as long as it is honestly entered into, and there are mechanisms for people to move and change their opinions as they are exposed to the views of others. This is very different from polling, which simply tries to learn what people already believe, and then calibrates arguments to reinforce it.

As Henry Farrell wrote to me in another email: "Processes of intellectual discovery are all about arguments between different (and sometimes stylized) positions. To use a machine learning analogy stolen from my collaborator, Cosma Shalizi—all of us put together are at best an ensemble of weak learners, each of which only grasps a few of the terms in a very long and complicated vector that we're trying to model. It plausibly helps if we start from very different positions (each weak learner sees a different set of terms) as long as each of these positions reflect some aspect of the truth and then, and only then, try to converge on a shared model of the problem."

That is a beautiful summation of the power of intellectual debate to drive toward truth. We face enormous challenges as a society as that debate moves into online platforms with billions of participants, with no boundaries of nationality or geography, with untested signals of authority and authenticity, using rude tools not yet up to the task.

It's still day one.

LONG-TERM TRUST AND THE MASTER ALGORITHM

Truth is only one of many factors humans—and the companies they create—struggle to optimize. What is really driving our decisions?

Some years ago, John Mattison, the chief medical information officer of Kaiser Permanente, the large integrated health provider, said to me, "The great question of the twenty-first century is going to be 'Whose black box do you trust?'" A black box, by definition, is a system whose inputs and outputs are known, but the process by which one is transformed to the other is unknown. Mattison was talking about the growing importance of algorithms in medicine, but his point, more broadly, was that we place our trust in systems whose methods for making decisions we do not understand.

Sometimes that trust is given because we ourselves don't have the knowledge to understand the algorithm, but we believe that someone else does. Sometimes that knowledge is denied even to experts capable of understanding what is inside the black box; it is kept from them as a trade secret. Google does not disclose the exact details of its search algorithm lest it be gamed by those trying to increase their rankings. Similarly, when Facebook cracked down on stories with clickbait headlines, Adam Mosseri, its VP of product management for News Feed, wrote, "Facebook won't be publicly publishing the multi-page document of guidelines for defining clickbait because 'a big part of this is actually spam, and if you expose exactly what we're doing and how we're doing it, they reverse engineer it and figure out how to get around it.'"

Just as with clickbait headlines, some incentives to create fake news can be eliminated. Many of those promoting fake news during the 2016 election were politically motivated, whether sincerely or cynically,

but many fake news sites, like the ones created by the Macedonian teens, were created purely for financial gain. Cutting off advertising for sites or accounts that are peddling fake news is a great way to eliminate some of the most egregious offenders. This can be done not only by the platforms themselves, but by advertisers and ad networks who place "remnant advertising" on the lowest-quality sites. Businesses are beginning to recognize that the ads they show against their content make a statement about who they are, and showing the wrong ads can irrevocably damage their own reputation. As Warren Buffett is reputed to have said, "It takes twenty years to build a reputation and five minutes to ruin it. If you think about that, you'll do things differently."

Outright bad actors are only a small part of the problem, though. A more fundamental challenge is the way that the fitness function in the algorithms of search and social media shape the choices made by writers and publishers. Advertising-driven businesses in particular are slaves to the need for attention. Chris O'Brien, formerly a reporter for the *San Jose Mercury News* and the *Los Angeles Times* and now at online publisher *VentureBeat*, told me of the struggle reporters like him face every day. Do they write and publish what they think is most newsworthy, or what will get the most attention on social media? Do they use the format that will do the most justice to the subject (a deep, authoritative piece of research, a so-called longread), or do they decide that it's more profitable to harvest attention with short, punchy articles, perhaps even with deceptive headlines, that generate higher views and more advertising dollars? Do they choose video over text, even when text would let them do a better job?

The need to get attention from search engines and social media is a major factor in the dumbing down of news media and a style of reporting that leads even great publications to a culture of hype, fake controversies, and other techniques to drive traffic. The race to the bottom has in part been a result of the primary shift of news industry revenue from subscription to advertising and from a secure base of local readers to chasing readers via social media.

Subscription-based publications have an incentive to serve their readers; advertising-based publications have an incentive to serve their

advertisers. As described in Chapter 8, search-based pay-per-click advertising can help to align the incentives, but it too can be gamed, and in any event it represents only half of digital ad spending, which in turn is only a fraction of total advertising spending. The flood of subscribers to news publications like the *New York Times*, *Washington Post*, and *Wall Street Journal* since the 2016 presidential election is a promising sign that there is interest from consumers in supporting investigative reporting again. But publications like these that formerly dominated the news media landscape are now much less influential. As a result, those whose algorithms guide what content is consumed via search and social media have a deep responsibility to tune their algorithms not just for profit but for the public interest.

Because many of the ad-based algorithms that shape our society are black boxes—either for reasons like those cited by Facebook's Adam Mosseri, or because they are, in the world of deep learning, inscrutable even to their creators—the question of trust is key. Facebook and Google tell us that their goals are laudable: to create a better user experience. But they are also businesses, and even creating a better user experience is intertwined with their other fitness function: making money.

Evan Williams has been struggling to find an answer to this problem. When he launched *Medium*, his follow-up to Twitter, in 2012, he wrote, rather presciently as it turned out: "The current system causes increasing amounts of misinformation . . . and pressure to put out more content more cheaply—depth, originality, or quality be damned. It's unsustainable and unsatisfying for producers and consumers alike. . . . We need a new model."

In January 2017, Ev realized that despite *Medium*'s success in building a community of writers who produce thoughtful content and a community of readers who value it, he had failed to find that new business model. He threw down the gauntlet, laid off a quarter of *Medium*'s staff, and committed to rethink everything it does. He had come to realize that however successful, *Medium* hadn't gone far enough in breaking with the past. He concluded that the broken system is ad-driven Internet media itself. "It simply doesn't serve people. In fact, it's

not designed to," he wrote. "The vast majority of articles, videos, and other 'content' we all consume on a daily basis is paid for—directly or indirectly—by corporations who are funding it in order to advance their goals. And it is measured, amplified, and rewarded based on its ability to do that. Period. As a result, we get . . . well, what we get. And it's getting worse."

Ev admits he doesn't know what the new model looks like, but he's convinced that it's essential to search for it. "To continue on this trajectory," he wrote, "put us at risk—even if we were successful, business-wise—of becoming an extension of a broken system."

It is very hard to repair that broken system without rebuilding trust. When the algorithms that reward the publishers and platforms are at variance with the algorithms that would benefit users, whose side do publishers come down on? Whose side do Google and Facebook come down on? Whose black box can we trust?

There's an irony here that everyone crying foul about the dangers of censorship in response to fake news should take deeply to heart. In 2014, Facebook's research group announced that it had run an experiment to see whether shifting the mix of stories that their readers saw could make people happy or sad. "In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed," the researchers wrote. "When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks."

The outcry was swift and severe. "To Facebook, we are all lab rats," trumpeted the *New York Times*.

Think about this for a moment. Virtually every consumer-facing Internet service uses constant experiments to make its service more addictive, to make content go viral, to increase its ad revenue or its e-commerce sales. Manipulation to make more money is taken for

granted, its techniques even taught and celebrated. But try to understand whether or not the posts that are shown influence people's emotional state? A disgraceful breach of research ethics!

There is a master algorithm that rules our society, and, with apologies to Pedro Domingos, it is not some powerful new approach to machine learning. It is a rule that was encoded into modern business decades ago, and has largely gone unchallenged since.

It is the algorithm that led CBS chairman Leslie Moonves to say in March 2016 that Trump's campaign "may not be good for America, but it's damn good for CBS."

You must please that algorithm if you want your business to thrive.