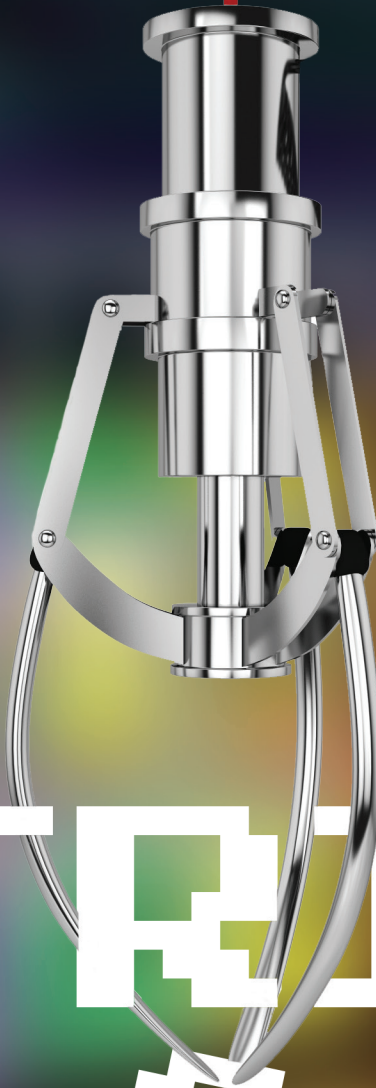


THE NEW



METRICS CRISIS

BY DAVID BERREBY

AS

companies lose confidence in ROI and other data they're getting, big tech firms and established marketers look for ways to regain trust.

The social-media campaign vowed to deliver 100,000 engagements. The client signed on. And now the question facing the agency, and the client's people who hired them, is pretty clear: How do we

make good on that promise? On paper, of course, it's a simple task to show precise ways to measure engagement. The reality is often different.

In fact, to hear Tim Hwang tell it, the math is pretty fuzzy—and he comes steeped in high-level experience in the tech industry. He is now director of the Harvard-MIT Ethics and Governance of AI Initiative. “You say, ‘OK, what are the things we need to add up in order to get to that number?’ And you might say, well, we know from Google Analytics that there were, like, 50,000 page views, right? And then maybe on top of that there’ve been, like, 20,000 clicks. So we’ll add that on,” he says.

Then, he adds, “Let’s say, you know, maybe 2 percent of those people decided to tell their friends about this. We don’t know that, but that’s probably a safe estimate.” Presto, more than 80,000 engagements are in the bag.

It’s no secret that when everyone involved has an incentive to show a campaign worked, they likely aren’t using metrics to find an answer. They’re using the answer to find their metrics. Indeed, a culture of fuzzy accounting around metrics that back the value of ads, social-media efforts, and other marketing campaigns—and determine a company’s true return on investment for all that—has grown as corporations increase their demands to know just what they’re paying for. For his part, Hwang likens it to all the mortgage industry in 2008: As millions of loans then turned out to be worth less than they were sold for, so may the value of people’s attention on the web be



THE PROBLEM

Firms are questioning data that show the value of ad and marketing content.

WHY IT MATTERS

Companies need reliable ways to sell and reach customers.

THE SOLUTION

Big tech firms and agencies need a full revamp of data collecting and set proper expectations.

worth less—a “subprime attention crisis,” as he fashions it.

None of this is a secret. In fact, the digital-media world regularly rings with calls for better metrics and more transparency. Last year, for example, when Simply Measured, an analytic software company, surveyed nearly 1,000 social-marketing specialists in ad agencies around the world to ask about their greatest challenges, 61 percent named “measuring ROI.”

But it isn’t only marketers, ad agencies, and consultants who have a problem. Even the big tech platforms through which users access the Internet—the companies that, unlike agencies and clients, actually have all the hard data on who clicks where and when—lately have had to revise, recall, or redefine metrics. Last summer, for example, many celebrities on Twitter suddenly lost hundreds of thousands of followers as the social network began purging itself of millions of fake accounts (many of them “bots,” created and maintained by software). President Donald Trump’s follower statistics dropped by 100,000 “people”; his predecessor, Barack Obama, lost 400,000 followers.

IN THE CHASE FOR THIS OR THAT NEW METRIC TREE, IT’S EASY TO LOSE SIGHT OF THE FOREST.

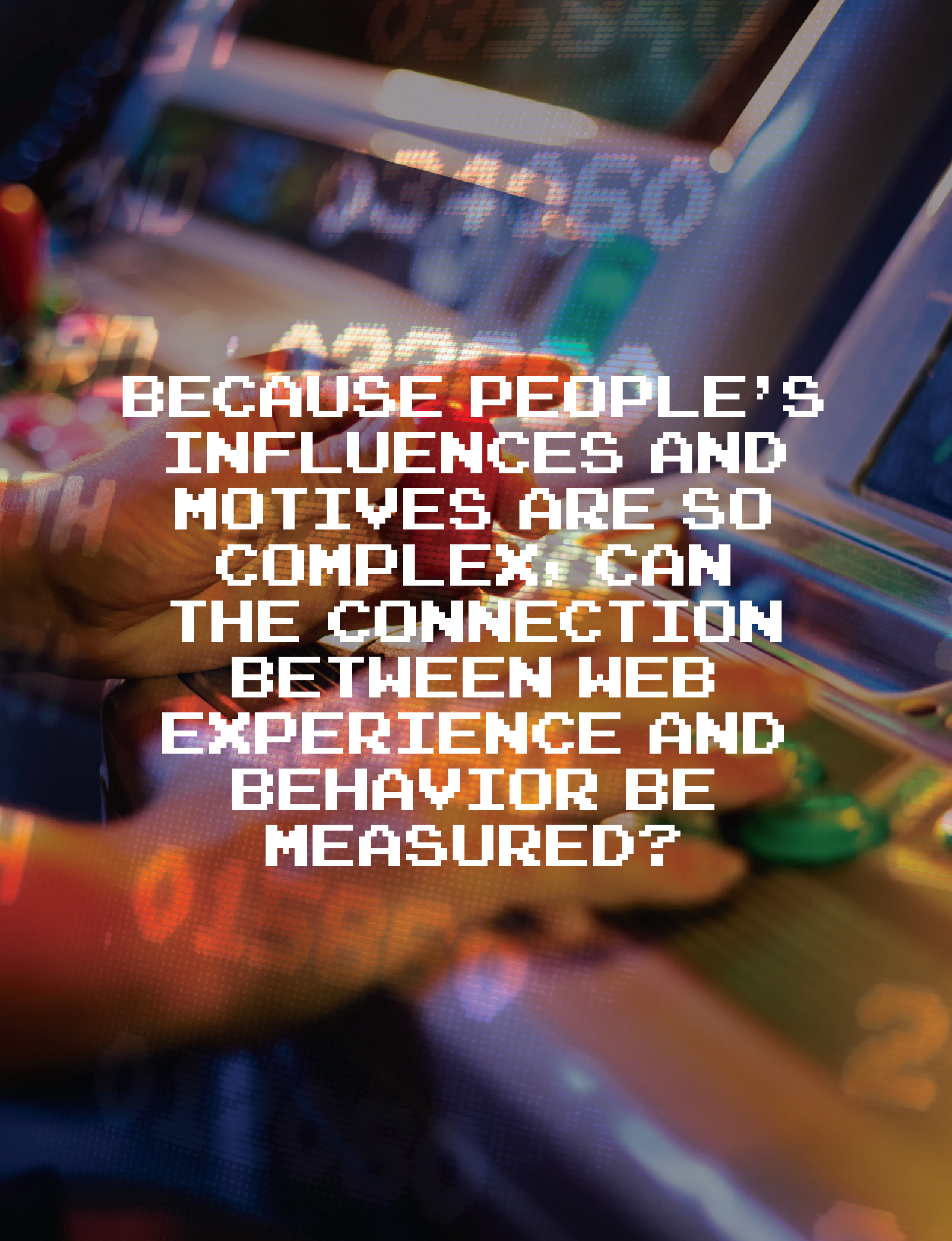
Around the same time, Instagram killed off a number of huge “comment pods”—groups with hundreds of thousands of members organized to like one another’s posts and thus boost their chances of being included in people’s feeds.

And then there was this: Last summer, AdNews Australia reported that Facebook’s Ad Manager tool was assuring users their ads would reach 1.7 million more Australians age 15-40 than actually exist. Shortly afterward, Brian Wieser, an ana-

lyst from the firm Pivotal Research Group, created a test ad and found Facebook’s Ad Manager claimed he could reach a potential audience of 41 million 18- to 24-year-olds in the United

States—a country that, according to the US Census Bureau, has only 31 million people in that age group. In response, Facebook removed that metric, and announced later it was revising about 20 metrics it deemed “unhelpful” in their current form.

Consultants say one reason for such flubs is that big tech firms are not single-minded monoliths. Different divisions, from product and engineering to sales and products, have different goals, from proving adoption to showing profitability.

A hand holding a pen is positioned over a document that features a grid pattern. The background is a complex, digital, and pixelated composition with various colors like blue, orange, and green, suggesting a high-tech or data-driven environment. The text is centered and written in a white, pixelated font.

BECAUSE PEOPLE'S
INFLUENCES AND
MOTIVES ARE SO
COMPLEX, CAN
THE CONNECTION
BETWEEN WEB
EXPERIENCE AND
BEHAVIOR BE
MEASURED?

Fledgling political scientists in the US are taught that Congress is “a ‘they,’ not an ‘it,’” notes K. Sabeel Rahman, a professor of law at Brooklyn Law School who is president of Demos, a “think-and-do tank” that focuses on political and economic equality. At a panel discussion last spring in New York City on tech platforms, he noted that a big technology company should be seen the same way. “We’re trained to remember that a legislature doesn’t act as a unit—it’s hundreds of people, organized in factions. It’s equally true that Facebook is a ‘they,’ not an ‘it,’ and that’s something we often forget.”

These different incentives can explain why tech companies haven’t been pointing to evidence that their influence on people is almost certainly not as strong as people fear. Such a message might please their just-the-facts engineers, and perhaps their lobbyists and public-policy staff, who reckon with public anger about “fake news” and “information bubbles.” But the argument would undermine the claim that salespeople make, that their advertising can change people’s behavior.

In any event, even optimists, who think metrics will eventually capture what they promise to capture, now recognize that the quest needs to focus less on counting likes, shares, retweets, or “engagements” and more on consequences that matter (for example, “conversions,” the industry’s term for the moment when the target of a message performs the action for which the message was bought, most typically buying some good or service). Last year, for example, Facebook COO Sheryl Sandberg told analysts that the company was going to move away from such “proxy metrics” to “sales metrics.” The reason, as she said, is that “the

more that we can tie ad viewing to sales, the stronger our case is with our clients.”

Institutional issues aren’t the only factor in the problem of fuzzy metrics. There are, of course, many technical challenges to measurement. For example, many data-gathering tools (including Google Analytics) store “cookies” of information on users’ computers in order to track them. But a growing number of people are refusing or deleting the cookies. (The ad serving company Flashtalking, in an analysis of 20 advertisers in the fourth quarter of 2017, found 64 percent of their tracking cookies were either blocked or deleted by web browsers. For mobile devices, the refusal rate was 75 percent.) Other issues—multiple devices used by one person, one device used by multiple people, the way location data in mobile devices is captured—are likely contributing to the fuzziness of current data.

Many keen minds are working on these technical challenges, for obvious reasons. Every incremental improvement in measurement yields some advantage to the organization that can offer it to the world. But in the chase for this or that new metric tree, it’s easy to lose sight of the forest. Beneath the issues of measurement are bigger questions about human behavior and how we understand it.

One recent study, by Matthew Gentzkow, a professor of economics at Stanford University, found that the actual influence of social media on most people’s politics was feeble. To believe that “fake news,” for example, had an effect on voting in 2016, he concluded, you would have to believe that a phony news story has the same persuasive effect as 36 television

WHAT CAN GO WRONG

Despite the involvement of hundreds of organizations and billions of dollars, here are some basic ways that metrics for digital marketing and advertising can go awry.

Organizational and individual inertia. It leads firms to use familiar instead of better measurements. One survey of ad agencies found that nearly 60 percent used engagement metrics to measure campaign effectiveness, while only a quarter used conversions, which more accurately tie digital behavior to sales.

ads. In the same vein, a recent study by the political scientists Andy Guess, Brendan Nyhan, and Jason Reifler found that while fake news was prevalent on Facebook, the people most likely to encounter and consume it were people who were already disposed to agree with it. Such results are one reason researchers in political behavior doubt that social media can change people's behavior. "Rather than only asking how Facebook affects opinion," wrote the political scientists Jessica Feezell and Yanna Krupnikov last year in *Behavioral Scientist*, "a better question may be to ask how people's opinions influence what they see on Facebook."

If that leaves you wondering why digital companies don't offer such points in defense when they are pilloried for their effects on society, remember that point about the many overlapping organizations that they contain. What is good politics can be bad salesmanship. "It really is a catch 22 for these companies," Hwang says. "Either it is influential and they're complicit in this problem, or it is not

influential and their business model's a fraud. That's a very difficult tight-rope to walk."

To some degree, a lot of today's metrics crisis is steeped in an age-old and fundamental problem: the difficulty of finding cause-and-effect chains in human behavior. In physical sciences like physics or chemistry, the bodies and forces involved in a causal chain are known and measured with extreme accuracy. The forces involved in human choices, though, are far more varied and numerous. They include passing moods, personality traits, memories, fears, what happened five minutes ago, as well as the rising and falling influence of family, friends, culture, money, and politics. This is why it's much easier to prove that Saturn's far-off moon Enceladus has an ocean full of water than it is to prove that a particular set of pixels seen at 11:23 a.m. yesterday caused someone to buy a product or sign up for a newsletter.

Perhaps it isn't an impossible goal

Measuring the wrong thing. For media companies, finding an audience that's willing to pay has been the key to success. Metrics that use clicks, shares, or sheer traffic can't capture this, but can make a small successful business look less promising than a company that has never made a profit.

Measurement that's too powerful. Some metrics do work, and that's the problem: Their technical achievements lead to undesirable consequences among consumers worried about tech-firm overreach and privacy.

Measuring where we should not. In nascent sectors, the lack of metrics has provided breathing room for innovation. Podcasts are one example where good numbers were hard to find, prompting more experimentation.

if you're willing to be extremely intrusive. A decade ago, for instance, Sony patented a smart TV feature that would stop playing a commercial if the viewer stood up and said the name of the brand in the ad (US Patent 8246454 B2). But in the real world, people don't want to feel surveilled and coerced. (Sony's stand-up-and-shout feature was widely ridiculed.) That leads some to wonder if the holy grail of metrics—proving the influence of a digital experience on actual behavior—is even possible.

Because people's influences and motives are so complex, there is no practical way, short of a 24/7 surveillance regime, to measure the connection between web experience and behavior. Instead, as Sandberg said, the industry uses "proxies"—for example, clicking yes to receive a newsletter is a proxy for interest in its subject. Repeated visits to a travel site are a proxy for being interested in taking a trip.

This means any digital metric should face this question: Is the measured thing a good proxy for the

thing we want to know about? The difficulty of answering that question with certainty is the deepest and most stubborn reason why there's a gap between metrics hype and metrics reality. There is always a risk that—like the drunk in the proverbial story, looking for his lost keys by a lamppost because that's where the light is—researchers are measuring something because it can be measured, rather than because it's a good indicator.

For a question like "can we get this guy to buy something?" Hwang says, "Our metrics are still very poor proxies. The thing you're trying to measure doesn't really manifest itself through the screen." The standard tech-industry response, he says, "has been 'OK, we'll collect more data.'"

It may be, he adds, that steadily increasing amounts of data are getting the industry closer to an actual metric of consumer interest. But it may be instead "that actually it's very difficult to get what we need to know from a screen." ▀