



National Endowment
for Democracy
Supporting freedom around the world

forum
International Forum for Democratic Studies

FORUM Q&A: KELLY BORN ON THE EVOLVING FIELD OF DISINFORMATION RESEARCH

As executive director of the Cyber Policy Center, Kelly Born collaborates with the center’s program leaders to pioneer academic programs focused on cyber issues, including new lines of research, a case-based, policy-oriented curriculum, pre- and postdoctoral training and practitioner fellowships, policy workshops and executive education. Prior to joining Stanford, Born helped to launch and lead the U.S. Democracy team at the William and Flora Hewlett Foundation, one of the largest philanthropic undertakings in America working to reduce polarization and improve U.S. democracy. There, Born designed and implemented strategies focused on money in politics, electoral reform, civic engagement and digital disinformation. In this latter capacity, Born worked with academics, government leaders, social media companies, foundations, and nonprofits around the world to help improve online information ecosystems.

Disinformation—the purposeful dissemination of misleading content to divide audiences, undermine social cohesion, or some combination of the two—has in recent years become a leading concern for political observers following its role in high-profile elections around the world. The use of disinformation is not new, but the space in which it is employed by political actors has changed swiftly and radically. A growing community of researchers is debating the consequences of these changes for democracy, and how societies should respond.

Dean Jackson of the International Forum for Democratic Studies spoke with Kelly Born about how the field of disinformation research has grown and where it should go from here.

Dean Jackson: Kelly, in 2016, you were one of the earliest funders of research into disinformation and its political impact. What has changed in this space since then?

Kelly Born: One obvious thing that’s changed since 2016 is the “techlash.” The Arab Spring and the publication of my colleague Larry Diamond’s article on “[Liberation Technology](#)” were only a few years before 2016. We were still in a state of techno-utopia—recognizing the many benefits that social technologies can provide, but without yet appreciating how they could be abused by bad actors, or how the business models of these platforms might prove problematic. We are now in a dramatically different place.

A second change is that there is now a field of people studying disinformation, with academic centers, think tanks, listservs, and a real community of experts. We had almost none of that connective tissue before 2016.

We also have frameworks to organize what we’re seeing. In the early days, [I wrote about](#) what I saw as the three main points of intervention: “upstream”, working to improve the quality of journalism; downstream, on citizen-facing efforts like fact-checking and news literacy; or (much less common at the time) mid-stream, working to improve the distribution of content by new online platforms. It is this latter area, of how information is

distributed, that has changed most dramatically in the modern era, and where interventions are more readily scaled. But it needed to be better understood.

And we now have frameworks for thinking about how to detect and address problematic content. In the early days, the conversation was about whether content was true or false. But often, the kind of content we are most concerned about is not categorically true or false—it is heavily biased, misleading, or inflammatory. We now realize that in addition to looking at the content itself, it's helpful to think about the actor behind the content, or the behaviors or techniques that actor is employing to amplify their content—creating fake accounts, running bot networks, or micro targeting in a discriminatory way. Camille Francoise recently summarized this as the “ABCs of disinformation—“[actors, behaviors, and content.](#)”

Finally, we also have a much more nuanced idea of what platforms can do about problematic content. Initially the thinking was that platforms should delete it, which of course runs into free-speech complications—especially in the United States where we have a much more absolutist view of free speech than [anywhere else in the world.](#) We now realize that in addition to deleting content, platforms can demote it, disclose the source, delay content that has reached a certain threshold of virality until it's verified, dilute it amidst higher quality content, deter (profit motivated) actors from placing it, or offer digital literacy, etc. My colleague Nate Persily framed this as the “[7 Ds of Disinformation.](#)”

And, in addition to frameworks, we now have a better understanding of exactly how platform business models contribute to the problems we are experiencing. There was a great [MIT study](#) in 2018 which essentially found that false news travels six times faster than the truth. That's a direct function of how these types of socio-technical systems work: People like stuff that's novel and emotionally engaging. We also know that the business models of Facebook, Google, and Twitter rely on user engagement to sell ads, and thus that disinformation can drive engagement. We have learned so much in the last ten years.

One last thing: We realized early on that we didn't have the data we needed to answer questions essential to globally democracy. Now is the first time in human history that private companies hold such a rich trove of data about users, and generally speaking they're not sharing—both for legitimate privacy reasons, and also perhaps to shield themselves from liability. Greater transparency is necessary, but not sufficient; it will be an important precursor to sensible reforms.

You helped create [Social Science One](#), which was intended to help scholars gain access to privacy-protected data to help understand how social media platforms are impacting elections, and democracies, globally. How has researcher access to data changed or improved in the past few years?

I think everyone underestimated how hard it would be legally, given privacy concerns around sensitive personal data, and how difficult it would be operationally and logistically. Where do you even store the type and amount of data we were talking about? Do you put it on a clean, protected laptop that can't be connected to the internet? What if somebody steals the laptop?

To address privacy challenges, we ended up pursuing a new approach called differential privacy. The raw data collected by these companies are so detailed it would be easy to unmask a given user, even after the data was anonymized. Differential privacy involves introducing enough “noise,” essentially incorrect information about a user, that they can't be identified, but not so much noise that the data becomes useless for statistical analyses.

Eventually the difficulties were so great, and the time horizons for this first project so unclear, that the initial coalition of funders chose to exit the partnership. But the work continues. Something like 120 scholars from across 25 or 30 schools have received Facebook data. So there has been some improvement. But Google, YouTube, and Twitter never entered that partnership. Significant data access challenges persist with all platforms.

The story of Social Science One nicely illustrates one of the many values trade-offs we wrestle with in governing modern technologies: We want privacy *and* we want transparency and accountability. If you want perfect privacy, no one should ever see that data—not even Facebook. If you want perfect accountability and transparency, everyone and their mother should be able to look under the hood.

In a similar way, we want both free speech and an accurately informed public. We want a diversity of viewpoints, but we want users to primarily consume high-quality content. There are myriad, fundamental tensions in democratic values when it comes to the governance of digital technologies, and I think as a society we need to determine where on the spectrum we want to fall for each of these values trade-offs.

What are the most important outstanding questions that researchers could address with more data access?

After 2016, researchers were busy trying to [learn several things](#): how much disinformation was out there, who was sharing it, who was consuming it, how much of it was organic versus driven by the Facebook algorithm, and did it change people's beliefs and behaviors or was it just preaching to the choir? We have learned a great deal since then—much of which is documented in an edited volume we just published with colleagues at New York University's Center on Social Media and Politics, a book called [Social Media and Democracy: The State of the Field, Prospects for Reform](#).

I believe the most important questions now are around impact, and solutions.

We still know very little about the actual impact of online disinformation on people's beliefs and behaviors. It's not just "fake news" that we're concerned about (and many have taken issues with how narrowly that is defined). It's also content that is biased, polarizing, or inflammatory. And there are impacts beyond vote choice, including questions about the effect social media is having on voter demobilization, racial divisions, political polarization, misogyny, and even declining trust in media and other institutions.

While outside researchers have yet to definitively answer some of these questions about impact, a [Wall Street Journal expose](#) revealed internal Facebook reports from 2016 which found that "our algorithms exploit the human brain's attraction to divisiveness," and another internal report from 2018 noting that 64 percent of people who joined an extremist group on Facebook only did so because the company's algorithm recommended it to them.

With more data, and a clearly defined, impact-oriented research agenda, much more could be learned about the various ways social media are impacting global societies.

Second, looking at solutions, the place to begin, I believe, is with ongoing improvements to transparency. We need to be able to continuously re-examine the impact these changes are having on our societies.

Arguably the platforms are transparent in some respects, as the transparency reports of [Facebook](#), [Google](#), and [Twitter](#) illustrate. These include, for example, the number of government takedown requests they accede to, archives of the political ads that have been

posted on their sites, or the numbers of videos removed year over year. They even include links to [“Transparency Reports on Influence Operations”](#) (but there are no actual data available).

What you soon realize is: Nowhere in all of these reports do the platforms share information about the role they themselves, their algorithms, are playing as independent actors heavily influencing what content people see and don't see. Everything is about what users, governments, political advertisers, etc. are doing on their platforms, but not how the platforms *themselves* act as independent variables of influence.

Even with Social Science One, the data Facebook agreed to share included all URLs shared more than 100 times on Facebook anywhere in the world, the socio-demographic data of those who shared it, and whether it has been subject to fact-checking. Even that fell short of the holy grail researchers had hoped for, as it still failed to get at the fundamental question of how the platforms themselves are influencing our societies. We need to know more about this.

In short, a few years ago we needed basic descriptive statistics of what was happening. We have more of that now. What we need today is to better understand the impact that this content is having, especially on disadvantaged communities, and how the platforms themselves—above and beyond what individual users are posting—might be driving audiences to that problematic content.

In the absence of data that could tell us for sure, how do you believe platform algorithms drive the spread of mis- and disinformation?

Most of these are digital advertising business models that prioritize engagement. As noted earlier, we know from prior research that people tend to engage more with emotionally stimulating content. Choices like prominently displaying number of followers, number of likes, and number of shares privileges popularity over other potential metrics, like quality.

And these social media sites are more than just one-way information sources. Sites like Twitter are also a key source of leads for many journalists, illustrating how platforms can help spread disinformation both directly to their audiences, and indirectly via their influence on what mainstream media covers.

And, for perhaps the first time, these are also social organizing platforms in a way that mainstream media never have been. Maybe you could advertise a protest in the newspaper, but the content wasn't as inflammatory, and the time lag for back-and-forth communication between participants was massive. With social media the ability to move from an online conversation to offline action is greatly enhanced. We saw this as an asset under the Arab Spring, but it proves deeply problematic when people are organizing towards less noble ends.

What are the most important next steps to better connect the many initiatives currently working on these challenges?

One thing I am trying to build out at Stanford is a better communications infrastructure around the research, to make sure we go that final mile in translating detailed research findings into clear policy recommendations. We are launching a monthly Policy Briefing series summarizing key findings from our research at Stanford for platform and government policymakers, as well as a weekly Tech Policy Watch tracking news in the field. We're also

hosting biweekly webinars with leading experts from around the world focused on these topics. In addition to the research itself, a key priority now is translating it to policymakers, both in government and tech companies.

The pandemic has revealed more information about tech platforms, but it is also accelerating the contest between illiberal and democratic norms in cyberspace more generally. How do you see this contest of ideas changing, post-COVID?

Digital technologies threaten to affect the balance of power both within existing democracies and globally, between authoritarian and democratic regimes.

Even within democratic societies, we see [landlords using algorithms](#) to screen tenants, police using facial recognition technologies to identify suspects, judges using algorithms to inform [prison sentencing](#), and employers using [gender-biased hiring algorithms](#). This reliance on potentially biased algorithms threatens to further harm groups that are already underprivileged or exploited, in ways that are inherently undemocratic. As COVID encourages more and more reliance on digital technologies, these risks increase if not addressed.

Now, under COVID-19, we are seeing unprecedented data collection about citizens' health and geolocation (the latter for contact tracing) around the world. This raises real concerns about government surveillance that may extend far beyond the pandemic—especially in authoritarian regimes with fewer privacy safeguards. Without sunset clauses or other democratic restraints, these new surveillance capacities pose a wide range of threats.

And of course data is *the* critical input for improving AI capabilities. Increasing surveillance in countries like China both provide the government with more direct control, while also indirectly empowering them by strengthening their [AI capabilities](#), offering them a real advantage over democratic societies with less data access.

We are already well into a [global democratic recession](#). If China wins the AI arms race, while the U.S. continues to experience democratic gridlock and indecision, what will that mean for the fate of global democracies? The future balance of power *between* countries, and the extent of fairness or exploitation *within* our democracies, is going to be heavily influenced by digital systems

What does success in confronting these challenges look like?

I don't think there is going to be a silver bullet. In terms of solutions, it's going to be fifty different things, each of which get at a [small percentage of the problem](#). It's a never-ending arms race, but I believe it can get better. Looking at the problem of disinformation specifically, I think what "better" would look like, in a very abstract way, is a higher quality online information ecosystem.

Talia Stroud at the University of Texas did some [great research](#) back in 2013 when we first started this work—looking at what would happen if you replaced a "like" button with a "respect" button. You get very different outcomes.

Ultimately, I want to see an information ecosystem where the highest quality information is the information that is most widely shared. What floats to the top? Right now, by any measure, it is not the highest quality information. It's the inflammatory stuff that maximizes

engagement. This is a choice. It is a choice to let algorithms promote the most incendiary content. The internet doesn't have to look this way.

This interview has been condensed and edited for clarity. The views and opinions expressed here do not necessarily reflect those of the National Endowment for Democracy.