Information Services & Use 40 (2020) 259–272
DOI 10.3233/ISU-200098
IOS Press

Questioning the legitimacy of data

danah boyd*
Partner Researcher, Microsoft Research, New York, NY, USA

Abstract. This paper is based upon the closing keynote presentation that was given by danah boyd at the inaugural NISO Plus conference held from February 23–25, 2020 in Baltimore, MD (USA). It focuses on how data are used, and how they can be manipulated to meet specific objectives – both good and bad. The paper reinforces the importance of understanding the biases and limitations of any data set. Topics covered include data quality, data voids, data infrastructures, alternative facts, and agnotology. The paper stresses that data become legitimate because we collectively believe that those data are sound, valid, and fit for use. This not only means that there is power in collecting and disseminating data, but also that there is power in interpreting and manipulating the data. The struggle over data's legitimacy says more about our society – and our values – than it says about the data itself.

Keywords: Data legitimacy, data quality, data bias, data voids, agnotology, sociotechnical challenges, problematic queries, alternative facts, data infrastructure

1. Introduction

This paper is about questioning the legitimacy of data. We often talk about data as though it is a natural resource such as oil – one that fuels opportunity, but also one that has the potential to devastate the planet via oil spills. We talk about data with reverence, as if increasing the amount of data would give us the material with which to solve all of the world's problems. That hype is really ever-present, especially as the data conversation has been turning more and more towards Artificial Intelligence (AI) in every context.

I have often struggled with what people mean when they talk about AI. Since I come from computer science, I have a specific technical definition for the term. But when I look at the ads that surround us at airports, I know that the conversation about AI in business is not the same conversation about AI that I have experienced in technical circles. I was trying to puzzle my way through the expansive use of this term so I asked a tech executive: why is everyone talking about AI? He argued that we talk about "artificial intelligence" because we can't talk about natural stupidity.

The conversation about AI has little to do with the actual technologies at play. What surrounds us at the airport is a set of processes, myths, and ideologies. And that brings me to a quote by Geoff Bowker: "raw data is both an oxymoron and a bad idea. Data should be cooked with care" [1]. Within the context of AI we need to talk about what that data is, what it looks like, where it comes from, and what the nuances are. We need to tease out these issues in a sensible way so that we can better understand what makes data legitimate.

*E-mail: dmb@microsoft.com.

0167-5265/20/$35.00 © 2020 – IOS Press and the authors.
This article is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC 4.0).

Data has a serious problem: the moment it has significant power, people will try to mess with it. But if you think about it, this is not new. There is an old saying in economics: "If you torture the data long enough, they will confess to anything" [2]. As a social scientist, I know that data do not speak for themselves. Data are used for all sorts of purposes to say all sorts of things. If you want to understand data, you need to understand the systems in which they are embedded; this will reveal how that data can be contorted. The more that we are collectively obsessed with the importance of data, the more that data becomes vulnerable to all sorts of attacks. And when data are vulnerable, they are rendered in ways that serve a range of purposes.

Jeff Hammerbacher was the founding data scientist at Facebook. He left the company pretty publicly when he decided that he could not watch the best minds of his generation spend their time making ads [3]. He gave a talk at Data & Society [4]. and people asked him about how to make sense of what was happening in Facebook. He stated the following: "I found that the people who ascribe the most power to statistics and data are not people who do statistics and data science. They are executives who give the vision talks about the power of data. … I've seen so many cringe-inducing assertions … In my head, I'm listening to all these things and am like, I remember that conversation, and the data on which that is based is so utterly flawed and unlikely to be true. But it supports the mythos of this particular executive. So let's just repeat it until it is true" [5].

As I reflect on this quote, all I can think about are the AI advertisements that exist in airports. They both share a recognition that data is political, that data can be used to serve powerful people's interests by enabling illusions [6]. While this may appear to be a product of so-called "big data", there is nothing new about this phenomenon. Jill Lepore's account of the Simulmatics Corporation highlights how the mirage of AI and data analytics were at the center of the 1960 U.S. election and the Vietnam War [7]. Data tell stories. And they can be used to sell lies.

The legitimacy of data is not simply about the data's quality or its purported accuracy. Data's legitimacy comes from a belief that we can collectively believe that those data are sound, valid, and fit for use. This belief can be contested on a regular basis. There are times when we label data as being legitimate, but its quality is problematic, and there are times when the data appear to be of high quality, but its legitimacy is challenged. So I want you to hold both of these factors in your mind – quality and legitimacy – as I discuss different aspects of how we get to a data ecosystem, and what needs to be done when we have it.

2. Data and its bias

On my first day of teaching an introduction to data science class, I asked students to spend time working with data to show that they had invested in setting up their programming environment. I posted a URL on a screen and asked them to load a specific file: the New York City Stop and Frisk data set. To show that they successfully read in the file, I asked them to tell me the average age of someone who has been stopped.

Hand after hand comes up, with all students eventually arriving at the same answer: the average age was twenty-seven. I asked if that number was accurate, and everyone looked at me with furrowed brows: "what do you mean is it accurate?" From their vantage point, they had all arrived at the same answer so, clearly, what they did was right. I asked them what this number meant. They started projecting their own ideas, values, and social norms onto the data. One said they said that they thought that the average age was going to be lower, but perhaps there are a lot of homeless people getting stopped which must skew the age. I then asked: is the data right? Again, the students stared at me, unsure of what I could possibly mean.

I asked the students to look at the list of variables, noting that there was both a variable for age and one for date of birth. I asked them to see if these data matched. They did not. I then asked them to run a distribution of the ages. This produced an audible gasp as they realized that a sizable number of people in the data were either 99 or 0. I asked if they thought that this many newborns and near-centenarians had been arrested, and their newfound discomfort with the data became visible. The data were flawed, but the students had started running analyses without assessing the limitations or biases of the data.

Data is messy. It can be filled with all sorts of error. And that error can be introduced in many natural and unnatural ways. It can be introduced because our instruments are not calibrated correctly. It can be introduced by bias – by who is trying to do what with the data. It can be introduced by how the data are taken from one environment to another. As data scientists, we must first examine the limits of our data. It's crucial to understand how our data are made.

Latanya Sweeney [8] is a Professor of Government and Technology in Residence at Harvard University. One day, she was sitting down with a journalist and trying to recall a paper that she had written. So, she entered her name into Google, assuming that would be the fastest way to find the paper. While she was looking at the results, the journalist noted the advertisements that appeared adjacent to Latanya's name and asked why she might be getting so many ads for criminal-justice-related products. Latanya thought that this was odd and decided to run an experiment and test Google's system. She compiled a set of different baby names and looked at the racial categorization of those baby names. She used a subset of those names to search Google and quickly realized that one company was showing six different advertisements when people searched for names. Searches for Black-associated names frequently turned up ads that offered criminal background checks; this was not true for names that were primarily given to white babies, who most frequently got general background check ads with no connection to criminality.

She was intrigued. A casual searcher might immediately assume that the results demonstrated that Google was doing something nefarious by segmenting out baby names. But Latanya knew that this was not the case. Google offers the option of allocating its resources to a category such as names. Using machine learning, Google's ad system then works to provide the searcher with the most relevant ads (based upon the category selected by the advertiser) in order to give their client, the advertiser, the best return on their investment. In this case, when people were searching on Google for names more frequently associated with Black individuals, they were more likely to click on ads for criminal justice products than if they searched for white-associated names. In other words, a racist public taught Google how to be racist on its behalf so that Google then could amplify that to all of us. I don't believe that any one of those individuals who was clicking on those ads really thought through how their own prejudices were shaping their decisions of what to click on in a search, let alone teaching Google to reinforce these racist associations. Certainly Google, who is not categorizing these names by their racial histories, was not thinking that a racist public was going to create this feedback loop. But here we are. This is just one way in which societal biases can get baked into data-driven systems.

Virginia Eubanks [9], Associate Professor of Political Science at SUNY-Albany, has spent a great deal of time trying to understand how technical systems perpetuate longstanding prejudices even when their designers are working to combat prejudice. Machine learning systems and predictive technologies train on data from the past, often dooming us to repeat past mistakes. If the past is extremely problematic, so will be the predicted future. In her book, *Automating Inequality* [10], Virginia describes how even the best-designed systems by the most well-intended stakeholders can reproduce systemic problems once those technologies are placed into politically contested situations. As my colleagues and I noted in some

of our recent work, efforts at trying to build fair algorithmic systems are doomed if we don't account for the broader context [11].

To better understand these fraught dynamics, consider the role of algorithmic systems in the policing and criminal justice context. First, it's important to understand that both policing and criminal justice in the United States have a horrific racist history. As Michelle Alexander accounts for in *The New Jim Crow*, contemporary mass incarceration is best understood as an extension of the country's original sin: slavery [12]. The data used in predictive policing, risk assessment scoring, and other algorithmic systems in this space are rife with biases because they come from discriminatory practices.

Consider the problems with data used in predictive policing models [13]. Imagine building a model to identify criminal drug offenses. Surveys consistently show that white individuals are more likely to consume and sell drugs than Black individuals. But Black individuals are far more likely to be arrested and prosecuted for both the consumption and sale of drugs. The data available to law enforcement – from arrests to convictions – comes from the data that law enforcement collects, which is primarily about Black individuals. The data about whites is missing for many reasons. One such reason is that police do not routinely visit college campuses to look for drugs and make arrests, even though the drug consumption on college campuses is at the highest in any given geographic and demographic segment. (*Note*: Nearly one out of every four male college students and one in five female students uses drugs in one form or another [14].) Instead, police patrol low-income communities, detain people in those communities, and arrest people from those communities at a much higher rate. Thus the data they have suggests that drug activity happens in low-income communities. Fed back into a predictive policing system, the system will tell them to focus in on low-income communities. Many well-intended computer scientists want to "de-bias" data, to root out the longstanding racist histories, and reweight data accordingly. This does not work so well when the data are missing in the first place.

3. Humans in the loop

Another site of technological mediation in the criminal justice system concerns the rise of risk assessment scoring in judicial decision making [15]. The same well-intended computer scientists who want to de-bias policing also want to computationally identify racist judges and build models that create risks scores that are free from such bias. This, too, is naïve because it fails to account for the role that judges play in this system. Risk scores are provided to judges to help them determine whether to release someone. This process is not automated. Instead, it's one signal that judges receive during bail or sentencing. So what is the role of the score in practice?

Judges are employees; they are elected or appointed. Either way, they don't want to lose their jobs. So if judges feel as though a score contradicts their professional opinion, what do they do? More often than not, they lean towards the score if the score is more conservative. After all, they don't want to explain why somebody was let out on bail when the algorithm told them bail represented a risk. It is a lot easier to blame the algorithm. Yet, if the score suggested a low risk and the judge perceived a higher risk, they're more likely to take a more conservative approach and claim that the algorithm's recommendation was too risky.

Risk assessment scoring algorithms may be flawed, but with the human-in-the-loop, designers see the human as the backstop. After all, judges are supposed to be able to make the final decision. Yet, risk assessment scores have significant power; they sway judges. This sociotechnical system produces a gap.

Neither the human judge nor the algorithm can be held truly accountable. Yet, together, they produce outcomes that can have significant impact on human dignity and freedom.

Madeleine Elish [16], Program Director at Data & Society, was researching the history of autopilot in aviation. Much to my horror, she taught me that contemporary airplanes are not actually flown by pilots, but by an automated system known as autopilot. The human pilots are merely sitting in the cockpit to babysit machines. Between 1913 and the 1970s, most aircraft had a pilot, a co-pilot, and one or more navigators in the cockpit. During a series of Federal Aviation Administration hearings in the 1970s where regulators debated autopilot, the FAA concluded that a navigator was unnecessary, but mandated that the pilot and co-pilot must remain, just in case they needed to step in for the machine. In other words, the FAA thought that it was crucial to have a human-in-the-loop.

Today's pilots – or autopilot babysitters – don't get much practice flying airplanes. They have been systematically deskilled on the job. Yet, they are expected to step in when all goes wrong, identify a problem, and fix it. Unfortunately, when things are really going wrong with a system, it is not so easy for a human to jump in and quickly fix it, even if they are quite skilled and have had lot of practice (which today's pilots do not have). So, what ends up happening is that because a human is the last person to touch the machine, a plane crash is almost always declared an issue of human error rather than an issue of technical error, despite the fact that these accidents usually begin due to technical errors.

Those who dispute this fact like to point out what is termed the "miracle on the Hudson", the amazing landing of an Airbus A-320 on the Hudson river in New York City by Captain Chesley (Sully) Sullenberger III on January 15, 2009 [17]. But I would like to point out that while Sully worked as a commercial pilot (he no longer works in the profession), he trained people on weekends how to re-fly during emergencies. He was probably the most skilled person possible for making that successful landing on the Hudson. He was not a typical pilot because most pilots are not actually given the opportunity to train others on a regular basis.

Madeleine argues that pilots involved in accidents end up being a liability sponge. They absorb the liability of the technical system. She coined the phrase "moral crumple zone" to describe "how responsibility for an action may be misattributed to a human actor who had limited control over the behavior of an automated or autonomous system. Just as the crumple zone in a car is designed to absorb the force of impact in a crash, the human in a highly complex and automated system may become simply a component—accidentally or intentionally—that bears the brunt of the moral and legal responsibilities when the overall system malfunctions" [18]. The sociotechnical system is set up to absorb the pain on impact, but it is mostly the human who receives the brunt of the pain. The human is stuck in the middle. And, in the case of airplane crashes, the human is typically dead.

When Boeing 737 Max planes first started falling from the air, humans were once again blamed. It took multiple plane crashes until people started questioning whether the technical systems might have a problem. Indeed, what we would learn is that Boeing had designed the planes to make it impossible for humans to step in. And when they tried, the flawed technical system overrode them. To this day, those planes are still grounded because they cannot build a sociotechnical system with a functioning relationship between pilot and machine.

To meaningfully empower a human in the loop requires much more than asking them to step in when all goes wrong. It requires an alignment of interests, practices, and values. Judges cannot serve as a backstop to flawed algorithmic systems when their livelihoods are at stake. Pilots cannot serve as a backstop to broken autopilot systems when the hand-off between machine and human is designed to put the human in the worst position possible. Algorithmic systems have power and accounting for that power is a crucial part of designing a responsible system.

4. Data and power

Data also has power. To understand data's power requires understanding where the data comes from. While there are many kinds of data, much of the data that relates to social media and/or search engines implicates humans, their practices, and their activities. This data tells us about people. Data about people comes to the world of data science in one of three ways: by choice, circumstance, or coercion.

Most data scientists like to think that they are working with data offered freely to them, data by choice. Consider the early stages of a startup such as Fitbit. People were so excited back in 2007 when Fitbit came out because they could share their steps with family and friends through the service. People knew that Fitbit would have their data, but since Fitbit was providing a valuable service, people did not think too much about access to their personal data. Yet, as Fitbit grew, it started making data available to health insurance companies. Employers started demanding people wear Fitbits and report the data back to them. Fitbit stopped being about choice – and this changed how people interacted with Fitbit. Today, there are people who put their Fitbits on their dogs in order to create data performances.

Data by coercion is the other end of this spectrum. This is data that is extracted by people through the exertion of power. A good example of this is the "spit and acquit" program in Orange County, California where certain defendants charged with petty misdemeanors are offered a dismissal or plea deal in exchange for their DNA [19]. The Supreme Court, in a case titled Maryland versus King [20], ruled that collecting a DNA sample of people during an arrest is equivalent to collecting a fingerprint or a photograph. Unfortunately, the Court seems to have failed ninth grade biology because they argued that DNA is a unique identifier. (Hint: DNA provides information about networks of biologically-related people.)

When law enforcement officers request DNA in exchange for not being arrested, that is not a fair and consensual exchange; that is coercion. But what's at stake here is not just how this impacts the person who is given a false choice; this coercive collection of data also affects people who are not present, such as family members whose data are now sitting in police records. Imagine what it would be like to have a Los Angeles Police Officer knock on your door to ask you about your brother who is wanted for arrest when you are not even aware that you have a brother! That would be awkward. This is where we can start to see the ripple effect of the "spit and acquit" coercion. It is important to realize that coercion doesn't just harm the individual, but rather it harms the ecosystem as a whole. The ripple effects of coercive data practices are over time and across communities, and can result in countless unintended consequences.

While the dynamics around data choice and data coercion shape a lot of our policies, laws, and attitudes towards data, most data that we work with ends up in an analyst's hands because of circumstance. We give our data to Facebook because we want to be on Instagram with our friends. When we turn this data over, we don't want to have to think about how it might be used. Because of this ostrich-like behavior, company executives often argue that people don't care about privacy. Yet, they do; they just don't want to think about these issues. They simply hope that they can trust the data, that there are rules and regulations that prevent the data from abused. For most people, privacy is not about the control of data, but the control of a social situation [21].

How data are collected truly affects the analyses. Consider a context like medicine. Most people would love to support scientists in their pursuit of curing cancer. When precision medicine is explained to people, most are thrilled at the idea that they could receive better treatment. Yet, there is a long history of medical data being abused. Many Black individuals do not participate in scientific studies because of the long history of very coercive data practices.

Kadija Ferryman [22], Industry Assistant Professor at New York University, is a cultural anthropologist who studies the social, cultural, and ethical implications of health information technologies. Specifically, her research examines the impacts of health risk prediction technologies as they relate to marginalized groups. To illustrate how bias enters into the medical field, she references a memo that was sent out to doctors after a study showed that a particular biomarker was associated with a specific heart disease. The memo argued that patients should be screened and if the biomarker was found they should be flagged for this heart disease. The notice went out across all of New York and doctors began to notice that most of the patients who had the biomarker were Black. It turns out that the studies related to that biomarker and that heart disease did not account for who was in their sample and who was not. They ended-up with inaccurate science because of systemic issues related to data bias.

This is an example in an area where people are really trying to do their best and make a positive impact. But consider what's happening when the incentives are far more mixed. Scheduling software is regularly employed in retail stores, emergency rooms, and other contexts where workers' schedules are centrally managed. These systems are designed to read in relevant information and produce schedules that maximize for certain interests. Such systems could easily be designed to maximize the interests and desires of workers – to ensure that they are working with the people that they like the most, that their shifts are optimized for special needs such as child care, that they get the number of hours, days of the week, time of day that they want, etc. Someone would still have to do the night shift and there would need to be trade-offs, but workers could articulate for themselves what would be fair. Of course, that's not how most scheduling software works. These systems are designed to maximize the interests of employers and managers, who are incentivized to ensure that all shifts are covered as cheaply as possible. Employers are also incentivized to ensure that workers do not regularly work with the same people so that they cannot unionize. They are incentivized to ensure that workers do not have stable schedules so that they can be on call. And they are incentivized to ensure that workers only work a certain number of hours per week so that they do not have to pay costly benefits.

The constraints and optimization goals of scheduling software are defined technically by powerful actors who are seeking to optimize for their interests at the expense of the interests of their workers. The data that is used for optimization is not just the data provided by workers. Instead, as Karen Levy and Solon Barocas argue, employers often surveil both workers and customers to regulate workers; they call this refractive surveillance [23]. In this way, coercive and circumstantial data can be combined and deployed in ways that assert power over vulnerable people.

5. Vulnerability of the data infrastructure

When data have significant power, people will inevitably mess with those data in order to achieve their own interests. For large-scale data infrastructure – like that maintained by the U.S. Census Bureau – protecting and securing data is an ongoing challenge. After all, people have tried to mess with the U.S. census since 1790. The tactics and methods evolve, but manipulation is persistent.

Unfortunately, many people who work with data do not account for how their data can be manipulated, gamed, or undermined. While contemporary conversations concerning AI, big data, and data science emphasize the challenge of obtaining and cleaning data, they do not tend to focus on when and why data is intentionally manipulated to affect the system, the model, or the analysis. There are lots of vulnerable data infrastructures.

To understand how data can be manipulated, let me describe some of the work Michael Golebiewski, Principal Program Manager at Microsoft's Bing, and I have been doing regarding the vulnerabilities in

search engines. In 1998, while still graduate students at Stanford, Sergey Brin and Larry Page wrote a paper describing a prototype search engine called Google [24]. In describing their design goals, they explain the need to improve search quality, highlighting how existing search engines surface up too many "junk results". They saw PageRank, their algorithm, as the solution. They believed that they could get true and meaningful signal from the data out in the wild. Of course, once Google started operating in the wild, the search engine optimization (SEO) industry simply evolved to game the new ranking structures. Like Altavista and Infoseek before, Google had to respond to these efforts and evolve its approach accordingly. The game of cat-and-mouse continues to this day, even though companies such as Google have erected more and more barriers. The motivation for most SEO attacks is obvious – there is a lot of money to be made when your site appears at the top of a popular search query.

Search engines require a lot of data to be useful. When people search for common terms, there are millions of results that search engines rank to provide users with results. If you search for a term like "basketball", you will have tons of data made available to you, because there are so many web pages and other content that is about basketball. SEO on those pages is hard. But if you search for something that is much more esoteric there will be far fewer responses. This creates a certain vulnerability and people can exploit the lack of data in an attempt to shape search results in a way that can pull information seekers into nefarious environments. Michael and I refer to this vulnerability as "data voids" [25]. It's a lot easier to SEO results that people rarely search. Or to SEO terms associated with breaking news stories. In our work on "data voids", Michael Golebiewski and I describe many ways that this vulnerability is exploited, but in this paper, let me describe just one type of media manipulation that leverages data voids.

5.1. Data void: Conspiratorial terms

Shortly after the Sandy Hook, Connecticut elementary school shooting in 2012, a group of conspiracy theorists started hypothesizing that people who appeared on TV as witnesses or family members of dead children were actually actors working for the Obama Administration to build a case to challenge the 2nd Amendment. They started labeling these people as "crisis actors" [26]. This term had very little online content associated with it at the time, so conspiracy theorists went about building a web of content. They produced images showing that the same "actors" appeared after every shooting, worked on identifying people and associating them with the "deep state". They built a harassment network that targeted parents of deceased kids. One father died by suicide because he couldn't take the harassment. Other parents sued one of the conspiracy theorists who regularly propagated this message.

In a different world of media, this conspiracy theory would have been esoteric. But these media manipulators built a web of content online with significant SEO. And then, after each subsequent mass shooting, they targeted news media with this term, trying to get them to pick it up. After the Parkland school shooting in 2018, Anderson Cooper asked David Hogg if he was a crisis actor [27]. This produced the outcome that conspiracy theorists most desired. Thousands of people turned to search engines to search for the term "crisis actor" and they walked straight into a web of conspiracy.

This is not a new technique. In the 1990s, the political opeerative, Frank Luntz, was well-known for creating pithy phrases that would evoke reaction. He would design these terms using focus groups and then push them out to members of Congress so that they would repeat these terms repeatedly in front of the media. This created a drumbeat around concepts with deep political frames. Many of his terms are now widely recognizable: "climate change", "partial-birth abortion", "death tax", etc. Luntz was very effective at leveraging the media ecosystem of the 1990s to propagate strategic terms.

Twenty years later, this strategy is regularly deployed by conspiracy theorists, political operatives, media makers, and media manipulators. Yet, because the media ecosystem has changed, so too have the methods. Today, media manipulators work to ensure that there is a wide network of online content that grounds the term before they work to propel it into mainstream media coverage. The goal is to manipulate the data that search engines rely on. But, once again, we see the magnification of terms produced by this process. Some of the terms produced by conspiracy theorists that have gone mainstream include "deep state", "white genocide", "caravan", and "incel".

While we have little quantitative evidence about how effective these campaigns are at increasing the adoption of conspiracy theories, we do have case studies of real harm. The massacre of nine Black churchgoers at Emanuel African Methodist Episcopal Church in Charleston, South Carolina on June 17, 2015 was committed by a teenager whose manifesto detailed how his worldview was altered when he encountered a data void [28]. Trying to understand the Trayvon Martin shooting, this teenager went to Wikipedia where he encountered a strategically-staged term. After using Google to search the term, he was first exposed to a white nationalist website. He spent two years consuming increasingly more violent extremist content before entering the church, sitting among a group at a Bible study for over an hour, and then opening fire. He discussed his online trajectory as an awakening. At his trial, he had no remorse.

Strategic placement of terms is one way in which data voids are exploited. Breaking news stories, outdated terms, and problematic queries are also regularly targeted. In each of these types of attacks, the goal of a media manipulator is to leverage vulnerabilities in search engines to introduce people to a new way of seeing. While this type of media manipulation is used by advertisers and marketers to sell products, it is also used by extremists and conspiracy theorists to radicalize people and polarize society. This is a vulnerability in data infrastructure that cannot be easily patched by search engines, even as they try to shore up their resilience to such attacks.

6. Agnotology and manipulation

Data vulnerabilities – the ways in which our systems can be manipulated by altering the data ecosystem in order to achieve specifically nefarious objectives – can have serious consequences. When people abuse them, they can harm systems and people. Yet, to really understand why this approach is powerful, we must understand how ignorance is strategically manufactured.

Epistemology is the study of how we know what we know. Agnotology is the study of ignorance. Robert Proctor and Iain Boal coined this term in 1995, shortly before Proctor and Londa Schiebinger began analyzing the different types of ignorance. In their co-edited book "Agnotology" [29], they argue that there is ignorance that comes from not yet knowing, which is the domain of science. There's ignorance that comes from forgotten knowledge, such as that which is produced through the erasures of people through colonial oppression. But, there's a third type of ignorance: that which is strategically produced. Ignorance that is manufactured for political, financial, or ideological reasons. We live in a society shaped by agnotology.

Their edited volume includes many different case studies and analyses about how ignorance is manufactured, including a chapter by David Michaels on how the tobacco industry funded studies to seed doubt into the emerging consensus surrounding lung cancer. Another chapter written by Naomi Oreskes and Erik Conway would form the foundation of their popular book/film "Merchants of Doubt" [30], where they lay out evidence showing how our inability to reach consensus on climate change stems from strategically manufactured ignorance. As I write this, we are living through another case study of strategically manufactured ignorance with COVID-19. It's not simply that we don't have enough tests

to get a good baseline number; it's that there are politically interested parties invested in ensuring that the data are muddied. We are watching states manipulate death data or selectively choose what data to report. We are seeing data used to manufacture ignorance for political and economic reasons. This is why agnotology, or the study of ignorance, is so important.

The world of media manipulation is deeply entwined with the ongoing effort to undo knowledge, to destabilize information, to make it so that we don't know what we can trust. In 2017, Cory Doctorow described this politicized dynamic quite succinctly when he tried to explain "alternative facts" to his blog readers:

> We're not living through a crisis about what is true, we're living through a crisis about how we know whether something is true. We're not disagreeing about facts, we're disagreeing about epistemology. The "establishment" version of epistemology is, "We use evidence to arrive at the truth, vetted by independent verification (but trust us when we tell you that it's all been independently verified by people who were properly skeptical and not the bosom buddies of the people they were supposed to be fact-checking)".
>
> The "alternative facts" epistemological method goes like this: "The 'independent' experts who were supposed to be verifying the 'evidence-based' truth were actually in bed with the people they were supposed to be fact-checking. In the end, it's all a matter of faith, then: you either have faith that 'their' experts are being truthful, or you have faith that we are. Ask your gut, what version feels more truthful?" [31].

In addition to coining strategic terms, political propagandists have also systematically worked to destabilize knowledge. The strategic manufacturing of ignorance is a well-established information operations tactic. Consider the work of Russia Today (RT), a Russian-backed TV and "news" organization [32]. One of their advertising campaigns – known as the "Question More" campaign – was designed to attract viewers by destabilizing scientific and political consensus [33]. At first blush, this campaign appeared to serve as educational content, designed using the tenets of media literacy. For example, one ad read: "Is climate change more science fiction than science fact?" [34]. Underneath that headline was smaller-type text:

> "Just how reliable is the evidence that suggests that human activity impacts climate change? The answer isn't always clear-cut, but it's only possible to make a balanced judgment if you are better informed. By challenging the accepted view, we reveal the side of the news that you wouldn't normally see. Because we believe that the more you question, the more you know."

Now, if this were not about climate change, but about the production of evidence for a scientific consensus, we might believe that this type of question makes sense – in order to learn more you need to question and ask how information is being constructed. But RT's goal wasn't to encourage reasonable debate; it was to use false equivalency to undo consensus. Many of these ads were posted in London, where people were outraged. The city chose to prohibit the ads as propaganda. In response, RT constructed another advertising campaign about being censored in order to attract those who already distrusted the state.

Media manipulation is happening all around us, shaping the public's understandings of politics and science. Yet, science and politics are also entwined. Consider the rise of the anti-vaccination movement. This movement is birthed out of a conspiracy theory, based on fabricated evidence linking the MMR vaccination with autism. While the paper at the root of this controversy was retracted [35] and the results summarily debunked, an increasing number of people believe that vaccines are dangerous. What is more disturbing is that the more the Centers for Disease Control (CDC), well-known scientists, or the news media publish stories trying to calm the public by debunking this conspiracy, the more we see a "boomerang effect". In short, because people do not trust institutions – including the government, the

academy, and the news ecosystem – people also believe that public debunkings are designed to dupe the public and hide what's really happening.

Because of media manipulation, the boomerang effect causes additional damage. When people distrust institutions, they don't just doubt the information provided. They often take another step and "self-investigate" using search engines and alternative news sources. In doing so, they find content that is produced to affirm their conspiratorial thinking. Rather than being able to build scientific consensus, our media ecosystem is helping fragment knowledge. This is the foundation of political conspiracy theory groups such as QAnon [36] and the root of our inability to curb the COVID-19 pandemic.

7. Interpreting the information landscape

Francesca Tripodi is a sociologist and an Assistant Professor at the University of North Carolina, Chapel Hill [37]. She spent 2017–2018 embedded inside a community of Christian evangelicals in Virginia in order to better understand how this community made sense of the information landscape [38]. One night, after regularly participating in a weekly Bible study and witnessing the different scriptural inference steps that this group adopted, the pastor turned to analyze the tax reform bill that was under discussion in Congress. To Francesca's surprise, the pastor invited his congregants to analyze the bill using the same techniques used to analyze the Bible. By using scriptural inference, the group wasn't looking for facts; they were looking to understand the deeper meanings underlying the text.

Recognizing the power of scriptural inference helped her better understand how this community was using Google. On the day of Virginia's 2017 gubernatorial primary election, Francesca interviewed voters about how they got information about the candidates. Unlike general elections, primary elections require people to think beyond party politics and choose between multiple candidates in their preferred party. Voters consistently told her that they didn't trust "fake news" (by which they meant CNN); instead, they turned to Google. Francesca thought she understood what they meant, but after watching voters show her how they did their research, she realized she was wrong.

Rather than using Google to truly research candidates, these voters searched for each candidate's name and then scanned the results provided by Google. They didn't click on websites to go deeper. Instead, they assumed that Google was providing the full range of relevant information through its search results. They read the headlines, scanned the text, and used that to draw their own conclusions about the candidates. Francesca has said that she thinks that when these people went to Google, they thought that Google weighed facts instead of ranking results [39]. Employees at Google were flabbergasted by these results. This was not what they had designed the system to do and they had no idea from the signals that they analyzed that this might be how people might be using their system. Google employees expected people to use the results as a jumping off point, not the arbiter of truth.

8. Towards a more secure future

In a world where data and media are manipulated, where people draw on different epistemic frames, and where algorithmic systems reinforce structural inequities, how can we collectively imagine what a world shaped by evidence-based decision-making might look like? I am convinced that we are not going to fix the mess that we are in – certainly not in 2020 – by fact-checking news, or regulating social media, or de-biasing data. These are reasonable interventions, but they serve in many ways as Band-Aids because they

fail to recognize that we are operating in an environment that is structurally problematic – an environment in which information, trust, and the idea of what it means to know something, are all coming undone in critical ways. We need to view this as a sociotechnical problem. That is to say that the issues at play in this environment are neither purely technical nor purely social. The issues come from the entanglement of the two. In figuring out how to understand the social and the technical as truly entwined, we need to reconsider some of our assumptions and affirmatively recognize some of our goals.

To close this paper, I want to emphasize two moves forward. First, we must design all sociotechnical systems with an understanding and awareness that there will be adversaries trying to attack those systems. Second, we need to intentionally build social fabrics that can hold people and communities in thoughtful ways. So far, I have primarily focused on making visible some of the vulnerabilities, but let me take a moment to tease out the importance of a social fabric.

When I was a teenager strolling through the virtual communities of the mid-1990s, I was desperately lost. I was angry and angsty, naïve and vulnerable. I will never forget two people who I encountered online in those early days who sat with me and helped me make sense of all my feelings. It was a form of therapy, but those conversations also opened my eyes to see a bigger world than that of my small town. Many LGBTQ youth had similar experiences to me in those days – thoughtful, kind strangers who ensured they lived to see another day. Today, the internet is mainstream. Yet, the kind, thoughtful, grounding strangers of my youth did not scale. Instead, what I'm finding is that young people who express confusion on YouTube or Twitter, 4chan or Discord get no supportive feedback. Worse, a sizable number of them are targeted by those with hateful and extremist agendas. Like the young white nationalist in Charleston, they are radicalized by people who target vulnerable youth. He was offered support from strangers and a framework for thinking about the world by people who had an agenda.

Key to the social fabrics of society are the personal networks of individuals, what sociologists call "social networks". Radicalization and polarization are not produced simply by individuals interacting with information; they typically involve the reworking of social networks. When social networks are manipulated, the result is segregation – ideologically, experiential segregation.

In 1787, a group of revolutionaries gathered in what is known as the Constitutional Convention. George Washington, a renowned general, did not believe that he should speak so he stayed quiet. Only once did he decide to speak up; he argued that it would be dangerous for the nascent democracy if the House of Representatives were to represent forty thousand people. He said that a single representative could, at most, represent thirty thousand people; otherwise these representatives would not know their constituents and would not be able to hold the country together. Now we are at a point where that number is seven hundred and fifty thousand; most people do not know those who govern them. And most people do not trust the government. Fifty years ago, people knew journalists in their communities. Today, most people do not. And most people do not trust journalists. Generally speaking, Americans don't trust any institution if they do not know people who are a part of it – universities, the military, the tech industry. If you don't know a professional working in a sector, it is hard to trust the sector. Social networks matters; they are how we build trust.

As we look forward, we need to grapple with the sociotechnical challenges that we face. To do this, we need to understand that our social and data infrastructure are under attack, just as democracy is under attack. Our Founding Fathers never argued that our country was perfect. Instead, they argued that we would have to work hard to achieve a more perfect union. This is a never-ending effort, a commitment that we must make each and every day. To achieve a more perfect union, we need to work to repair our social fabric, grapple with sociotechnical vulnerabilities, and strengthen our data infrastructure.

About the Author

danah boyd is a Partner Researcher at Microsoft Research; the Founder and President of Data & Society; and a Visiting Professor at New York University. Her research is focused on addressing social and cultural inequities by understanding the relationship between technology and society. Her most recent books, "It's Complicated: The Social Lives of Networked Teens" and "Participatory Culture in a Networked Age", examine the intersection of everyday practices and social media. She is a 2011 Young Global Leader of the World Economic Forum; a member of the Council on Foreign Relations; a Director of both Crisis Text Line and Social Science Research Council; and a Trustee of the National Museum of the American Indian. She received a bachelor's degree in computer science from Brown University, a master's degree from the MIT Media Lab, and a Ph.D. in Information from the University of California, Berkeley. Her website is: https://www.danah.org/.

References

[1] G.C. Bowker, *Memory Practices in the Sciences,* MIT Press, 2005, ISBN 978-0-262-52489-6.
[2] R.H. Coase, *Essays on Economics and Economists,* University of Chicago Press, 1994, ISBN 0-226-11103-2.
[3] D. Baer, *Why Data God Jeffrey Hammerbacher left Facebook to found Cloudera,* Fast Company, 2013, see: https://www.fastcompany.com/3008436/why-data-god-jeffrey-hammerbacher-left-facebook-found-cloudera, accessed July 6, 2020.
[4] Data & Society, a non-profit research organization that studies the social implications of data-centric technologies and automation, founded by danah boyd, see: https://datasociety.net/, accessed July 4, 2020.
[5] J. Hammerbacher, *Databite no. 93,* Data & Society, https://www.youtube.com/watch?v=5hUFtOQ3OU8, accessed July 20, 2020.
[6] M.C. Elish and d. boyd, Situating methods in the magic of big data and artificial intelligence, *Communication Monographs* **85**(1) (2017), 57–80.
[7] J. Lepore, *If Then: How the Simulmatics Corporation Invented the Future,* Liveright, 2020.
[8] https://dataprivacylab.org/people/sweeney/.
[9] https://virginia-eubanks.com/.
[10] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor,* St. Martin's Press, 2018, ISBN 13 97811250074317.
[11] A.D. Selbst, d. boyd, S. Frielder, S. Venkatasubramanian and J. Vertesi, Fairness and abstraction in sociotechnical systems. in: *2019 ACM Conference on Fairness, Accountability, and Transparency (FAT*),* 2018, pp. 59–68. Available at SSRN: https://ssrn.com/abstract=3265913, accessed July 20, 2020.
[12] M. Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness,* The New Press, 2010.
[13] A primer on the topic can be found here: S. Brayne, A. Rosenblat and d. boyd, Predictive policing, in: *Data & Civil Rights Conference Primer*, 2015, http://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf, accessed July 20, 2020.
[14] Statistics of Street Drug Abuse on College Campuses – Concerning Findings, by the staff of the Recovery Village, see: https://www.addictionhope.com/opiates/statistics-of-street-drug-use-on-college-campuses-concerning-findings/, accessed July 4, 2020.
[15] A primer on the topic can be found here: A. Christin, A. Rosenblat and d. boyd, Courts and predictive algorithms, in: *Data & Civil Rights Conference Primer*, 2015, https://datasociety.net/wp-content/uploads/2015/10/Courts_and_Predictive_Algorithms.pdf, accessed July 20, 2020.
[16] https://datasociety.net/people/elish-madeleine-clare/, accessed August 16, 2020.
[17] A. St. John, What went right: Revisiting Captain "Sully" Sullenberger and the miracle on the Hudson, *Popular Mechanics* (2019), see: https://www.popularmechanics.com/flight/a4137/sully-sullenberger-us-air-flight-1549-miracle-hudson/, accessed July 6, 2020.
[18] M.C. Elish, Moral crumple zones: Cautionary tales in human-robot interactions, *Engaging Science, Technology, and Society* **5** (2019), see: https://estsjournal.org/index.php/ests/article/view/260, accessed July 6, 2020.
[19] A. Roth, Spit and acquit: Prosecutors as surveillance entrepreneurs, *California Law Review* **107** (2019), see: http://www.californialawreview.org/print/spit-and-acquit-prosecutors-as-surveillance-entrepreneurs/, accessed July 9, 2020.

[20] Maryland v. King, Legal Information Institute, Cornell Law School, see: https://www.law.cornell.edu/supremecourt/text/12-207, accessed July 13, 2020.
[21] A. Marwick and d. boyd, Networked privacy: How teenagers negotiate context in social media, *New Media & Society* **16**(7) (2014), 1051–1067.
[22] http://www.kadijaferryman.com/, accessed August 16, 2020.
[23] K. Levy and S. Barocas, Refractive surveillance: Monitoring customers to manage workers, *International Journal of Communication* **12** (2018), 23, Available at: https://ijoc.org/index.php/ijoc/article/view/7041. Date accessed: 10 Aug. 2020.
[24] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* **30**(1–7) (1998), 107–117.
[25] M. Golebiewski and d. boyd, *Data Voids: Where Missing Data can be Easily Exploited,* Data & Society, 2018, see: https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf, accessed July 13, 2020.
[26] R. Wiedman, The Sandy Hook Hoax, *New York Magazine* (2016), since updated, see: https://nymag.com/intelligencer/2016/09/the-sandy-hook-hoax.html, accessed July 11, 2020.
[27] Stoneman Douglas High School Shooting, Wikipedia, see: https://en.wikipedia.org/wiki/Stoneman_Douglas_High_School_shooting, accessed July 11, 2020.
[28] Charleston Church Shooting, Wikipedia, see: https://en.wikipedia.org/wiki/Charleston_church_shooting, accessed July 11, 2020.
[29] R.N. Proctor and L. Schiebinger, *Agnotology: The Making and Unmaking of Ignorance,* Stanford University Press, 2008, ISBN-10: 0804759014.
[30] Naomi Oreskes and Erik M. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming,* Bloomsbury Press, 2010.
[31] C. Doctorow, Three kinds of propaganda and what to do about them, *boingboing*, 2017, see: https://boingboing.net/2017/02/25/counternarratives-not-fact-che.html, accessed July 11, 2020.
[32] https://www.rt.com/.
[33] R. Coalson, Question More, Radio Free Europe/Radio Liberty, 2020, see: https://www.rferl.org/a/Question_More/1927299.html, accessed July 11, 2020.
[34] M.L. Richter, What We Know about RT (Russia Today), Kremlin Watch Report, 2017, p. 15, see: https://www.rferl.org/a/Question_More/1927299.html, accessed July 11, 2020.
[35] G. Harris, Journal retracts 1998 paper linking autism to vaccines, *The New York Times* (2010), see: https://bpi.edu/ourpages/auto/2011/9/6/46169267/AutismarticleTheNewYorkTimes.pdf, accessed July 11, 2020.
[36] QAnon, Wikipedia, see: https://en.wikipedia.org/wiki/QAnon, accessed August 17, 2020.
[37] https://ftripodi.com/, accessed August 16, 2020.
[38] F. Tripodi, *Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices,* Data & Society, 2018, see: https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Searching-for-Alternative-Facts.pdf, accessed July 13, 2020.
[39] L.H. Owen, From Bible Study to Google: How Some Christian Conservatives Fact-check the news and end up confirming their existing beliefs, Neiman Lab, May 23, 2018, see: https://www.niemanlab.org/2018/05/from-bible-study-to-google-how-some-christian-conservatives-fact-check-the-news-and-end-up-confirming-their-existing-beliefs/, accessed July 11, 2020.