Dear Reader:

As the 2020 presidential election approaches, election integrity has moved to the forefront of our national consciousness. We decided to confront this topic directly: On February 7, 2020, the Georgetown Law Technology Review hosted a symposium titled Election Integrity in the Networked Information Era. By engaging with thought leaders from across multiple disciplines, we hoped to explore how technological development has changed our electoral process and how we can maintain electoral integrity in the modern era.

Those technological developments have been extensive. Social media and the Internet have changed how voters receive and process information, and those same networks have also created new opportunities for voter suppression. Even our election infrastructure has changed. Technology is often held up as a way to increase democratic participation and voter turnout, but those benefits can only come to fruition if we understand and mitigate new technologies’ limitations.

We organized the symposium around four different panels, each featuring experts from a variety of disciplines. The majority of those experts contributed to this issue. The following speakers were present for each panel with the first speaker in each column serving as the moderator for the panel:

<table>
<thead>
<tr>
<th>Networked Media Ecologies &amp; Public Discourse</th>
<th>Voter Suppression &amp; Voter Turnout</th>
<th>Digital Technologies &amp; Voting</th>
<th>New Challenges in Election Oversight &amp; Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erin Carroll</td>
<td>Julie Cohen</td>
<td>Matt Blaze</td>
<td>Alexandra Givens</td>
</tr>
<tr>
<td>Mike Ananny</td>
<td>Cara Brumfield</td>
<td>Andrew Appel</td>
<td>Marc Lawrence-Apelbaum</td>
</tr>
<tr>
<td>Leticia Bode</td>
<td>Jessica Huseman</td>
<td>Barbara Simons</td>
<td>Patrick Day</td>
</tr>
<tr>
<td>Whitney Phillips</td>
<td>Kathryn Peters</td>
<td>Philip Stark</td>
<td>The Honorable Karen Kornbluh</td>
</tr>
<tr>
<td>Lam Vo</td>
<td>Nick Monaco</td>
<td>Kim Zetter</td>
<td>The Honorable Ellen Weintraub</td>
</tr>
</tbody>
</table>

This issue is organized by panel, and the introduction to each section includes a brief discussion of the panel and the goals behind it.
Additionally, we held a lunchtime poster presentation session to highlight the work that a number of Georgetown students have recently done in this area. Those students and their subjects included:

<table>
<thead>
<tr>
<th>Chris Conrad</th>
<th>Natalie Gideon</th>
<th>Lauren Renaud</th>
<th>Kendall Spencer &amp; Rachel Wehr</th>
</tr>
</thead>
</table>

We are also proud to publish Chris Conrad’s Student Note, from which his presentation was derived, in this issue and to say that we published Lauren Renaud’s work, which won our first annual writing contest, in Issue 4.1. We would like to extend a special thank you to all of our student presenters.

We concluded the symposium with a panel addressing election regulations, and, even though we do not think that we found any definite solutions, we do think that we have started a productive conversation about how to address the technological realities of our contemporary elections. We hope that you enjoy that conversation and take the views that we have collected here into consideration. Only through continued effort can we address the problems facing elections in the networked information era.

Finally, we close the issue with a series of Technology Explainers. These explainers are written to provide lawyers with an understanding of how various technologies work—they are explanations of technical concepts for the legally inclined. In addition to pieces by several of our Staff members, we are also excited to include a Technology Explainer by an outside author this semester. Gabriel Nicholas, a Research Fellow at the New York University School of Law Information Law Institute and the New York University Center for Cybersecurity, has written a piece titled Explaining Algorithmic Decisions, and we are grateful for his contribution.

Thank you for reading.

Sincerely,
The Editorial Board
Spring 2020
GENERAL INFORMATION

Subscriptions: The Georgetown Law Technology Review is primarily an online law review, and as such, GLTR encourages you to visit its website at https://georgetownlawtechreview.org/. There, you can find information about how to order print-on-demand copies of this and other issues of GLTR. You may also download GLTR’s material free of charge at that address.

Publication Timeline: As an online law review, GLTR publishes articles throughout the year to its website. Twice per year—once in the winter and once in the summer—the journal collects these articles into an issue format for print publication. Upon compilation of each print issue, GLTR will post information concerning the purchase of print issues.

Submissions: GLTR accepts unsolicited materials through multiple submissions platforms as well as the email address gltr-submissions@georgetown.edu. The most current submissions guidelines can be found on the GLTR website.

Copyright: All contents of this publication are copyrighted by the Georgetown Law Technology Review (GLTR) except where otherwise expressly indicated. For all pieces copyrighted by GLTR, GLTR permits copies to be made for classroom use as long as proper citations are provided and notification of use is given to GLTR by emailing gltr-editor@georgetown.edu. For those pieces where another copyright holder has been expressly indicated, classroom use likewise permitted provided that proper citations are used and the identified copyright holder is notified.

Contact: GLTR may be contacted by emailing gltr@georgetown.edu.
ADVISORY BOARD

Julie Cohen
MARK CLASTER MAMOLEN PROFESSOR OF LAW AND TECHNOLOGY
GEORGETOWN UNIVERSITY LAW CENTER

Alexandra Reeve Givens
FORMER EXECUTIVE DIRECTOR, 2016–2020
INSTITUTE FOR TECHNOLOGY LAW AND POLICY
GEORGETOWN UNIVERSITY LAW CENTER

Hillary Brill
INTERIM EXECUTIVE DIRECTOR
INSTITUTE FOR TECHNOLOGY LAW AND POLICY
GEORGETOWN UNIVERSITY LAW CENTER

Paul Ohm
ASSOCIATE DEAN FOR ACADEMIC AFFAIRS, PROFESSOR OF LAW
GEORGETOWN UNIVERSITY LAW CENTER

Tanina Rostain
PROFESSOR OF LAW
GEORGETOWN UNIVERSITY LAW CENTER
MASTHEAD

EDITOR-IN-CHIEF
Joshua Banker

MANAGING EDITOR
Harsimar Dhanoa

MANAGING EDITOR
Laura Ahmed

SENIOR ARTICLES EDITOR
Aaron Scheinman

SENIOR SOLICITATIONS EDITOR
Sarah Koslov

SENIOR NOTES EDITOR
Sherlyn Abdullah

SENIOR CASE COMMENTS EDITOR
Michael Rose

SENIOR TECHNOLOGY EXPLAINERS EDITOR
Jeffrey Brown

SENIOR LEGAL NEWS EDITOR
Temesgen Woldezion

DIRECTOR OF DEVELOPMENT
Nolan Fargo

DIRECTOR OF OUTREACH
Alexandra Coyle

DIRECTOR OF TECHNOLOGY
Avi Ginsberg

ASSISTANT EDITORS
John Black
Raymond Coscia
Jake DeBacher
Allison Elkman
Adam Gerchick

Kristen Logan
Shreya Kundur
Kelly McCluer
Alex Nealon
Daniel Passon
Peter Pyatigorsky

Alex Rhim
Leetal Weiss
Eric Westerhold
Ryan Whittington
Christina Wing
STAFF

Ryan Abercrombie
Joseph Baillargeon
Daniel Barabander
Zev Beeber
Richard Bernache
Evan Burroughs
Joseph Cahill
Daniel Carlen
Jordan Cohen
Caitlyn Cook
Laura Cummings
Drew Diedrich
Andrew Do
Joseph Ehrenkrantz
Corey Fitzpatrick
Nicole Fulk
Tyler Gerstein
Andres Gonzalez
Clinton Greub
Brittany N. Griffin
Samuel Hanks
Grace Harter
Isabella Havas
John Heflin
Rebecca Iafrati
Tyler Kaufman

Gabriel Khoury
Gabriela Larralde
Ladan Mohaddes
Jessica Monsell
Michelle Mount
Florence Noorinejad
Jenevieve Nutovits
Sofia Panero
Josh Pereira
Sam Pickerill
Molly Rosen
Jay Schuffenhauer
David Seidman
Shelby Smith
Kendall Spencer
Lyle Stewart
Priyanka Surapaneni
Haris Vrahliotis
Yangbeini Wang
Mary Weaver
Rachel Wehr
Ruiqiao Wen
Matthew Wells
May Yang
Jeff Liji Zhou
Yifan Zhu
# TABLE OF CONTENTS

## NETWORKED MEDIA ECOLOGIES & PUBLIC DISCOURSE

How We Talk About the Press ................................................................. 335
   *Erin C. Carroll*

Making Up Political People: How Social Media Create the Ideals, Definitions, and Probabilities of Political Speech ................................. 351
   *Mike Ananny*

User Correction as a Tool in the Battle Against Social Media Misinformation ................................................................. 367
   *Leticia Bode*

Light Disinfects .................................................................................. 379
   *Whitney Phillips*

How the Internet Created Multiple Publics ........................................... 399
   *Lam Thuy Vo*

## VOTER SUPPRESSION & VOTER TURNOUT

The Risks and Rewards of Conducting a Census in the Digital Age ............ 415
   *Cara Brumfield & Jae June Lee*

Technology to Invite, Inform, and Modernize: How Platforms Support Democratic Participation ............................ 429
   *Kathryn Peters*

Amplify the Party, Suppress the Opposition: Social Media, Bots, and Electoral Fraud ................................................................. 447
   *Samuel Woolley & Nicholas Monaco*

Student Note: The Pernicious Problem of Platform-Enabled Voter Intimidation ................................................................. 463
   *Christopher Conrad*

## DIGITAL TECHNOLOGIES & VOTING

Election Integrity and Technology: Vulnerabilities and Solutions ............... 505
   *Matt Blaze*
Evidence-Based Elections: Create a Meaningful Paper Trail, Then Audit 523  
Andrew W. Appel & Philip B. Stark

Why Internet Voting Is Dangerous ................................................................. 543  
Barbara Simons

NEW CHALLENGES IN ELECTION OVERSIGHT & REGULATION

Desperate Times Call for Desperate Measures: The Need for Federal Legislation to Address New Potential Election Day Threats .......... 567  
Marc Lawrence-Apfelbaum

Cambridge Analytica and Voter Privacy .......................................................... 583  
Patrick Day

Public Safety and Disinformation .................................................................... 609  
Karen Kornbluh & Eli Weiner

Section 230 ........................................................................................................ 625  
Ellen L. Weintraub & Thomas H. Moore

Tailoring Election Regulation: The Platform Is the Frame ............................ 641  
Julie E. Cohen

TECHNOLOGY EXPLAINERS

Cyberattacks and Election Integrity ................................................................. 665  
Daniel Barabander

Distributed Ledger Technology .......................................................................... 673  
Drew Diedrich

Bitcoin Off-Chain Transactions: Their Invention and Use .............................. 685  
Michelle Mount

Innovations in Internet Voting Systems ............................................................ 699  
Michelle Mount

Explaining Algorithmic Decisions .................................................................... 711  
Gabriel Nicholas
ELECTION INTEGRITY
IN THE
NETWORKED INFORMATION ERA
NETWORKED MEDIA ECOLOGIES & PUBLIC DISCOURSE

Taking inspiration from the way that social and new media has changed the election landscape, our first panel explored the intersection of networked intermediation with journalism and political discourse. Our focus was on which stories journalists will tell about the 2020 election and how they might tell them. We wanted to find out how technology might engender, further, or manipulate those stories as well as how technology might mitigate some of the issues that emerged from the 2016 election.

Our panelists came from a wide range of backgrounds. Erin Carroll, currently a Professor of Law, Legal Practice, at the Georgetown University Law Center and previously a journalist, moderated the panel. Joining Professor Carroll on stage were Mike Ananny of the University of Southern California, Leticia Bode of Georgetown University, Whitney Phillips of Syracuse University, and Lam Thuy Vo of BuzzFeed News. Together, these panelists represented views from legal, academic, and practical perspectives.

Each one of the panelists also contributed an article to this issue, and we are pleased to present them in the following pages.
HOW WE TALK ABOUT THE PRESS

Erin C. Carroll*

CITE AS: 4 GEO. L. TECH. REV. 335 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 335
II. THE POWER OF NAMING .................................................... 337
III. CALLING THE PRESS NAMES ........................................... 339
IV. WHY WHAT WE CALL THE PRESS MATTERS ....................... 344
V. DIVORCING FAKE FROM NEWS ......................................... 347
VI. CONCLUSION......................................................................... 349

I. INTRODUCTION

“I believe freedom begins with naming things.” Eve Ensler

When the American Dialect Society made “fake news” its 2017 “Word of the Year,” the press release suggested two related reasons for the organization’s decision. The first was the term’s ubiquity. The second was its slipperiness. Beyond referring to propaganda, the term fake news had become a weapon—a “rhetorical bludgeon” against the press.3

---

* Professor of Law, Legal Practice, Georgetown University Law Center. Many thanks to the Georgetown Law Technology Review, Georgetown Law’s Institute for Technology Law & Policy, and especially to Julie Cohen, Alexandra Givens, and Joshua Banker, for organizing the Election Integrity in the Networked Information Era symposium and inviting my participation. I am grateful also to the panelists for whom I was lucky to serve as moderator, Mike Ananny, Leticia Bode, Whitney Phillips, and Lam Thuy Vo. Their thinking has enriched my own. Many thanks as well to Amanda Levendowski for her thoughtful feedback on a draft.


3 Id. (quoting Ben Zimmer, chair of the American Dialect Society’s New Words Committee and a Wall Street Journal language columnist).


The continued and often-uncritical use of fake news should worry us. As thinkers across disciplines have recognized for centuries, what we name things matters.\footnote{See infra Part I.} It shapes the very way we understand these things. This phenomenon is especially true when it comes to naming the press.

Although conventional wisdom is that press power and freedom spring primarily from the First Amendment, in reality First Amendment doctrine is that the press has no greater rights than any other speaker.\footnote{See infra Part III.} Rather, the press’s power and freedom are derived in large part from customs and norms. And those customs and norms draw sustenance from the language of the courts, other institutions, and the public’s belief that the press serves the democratic functions of truthful educator, trusted proxy, and fair watchdog.

Press power is, in great part, rhetorical power.
This rhetorical power is especially fragile in our networked information sphere, where content is infinite and gatekeepers are few. As we spend more of our lives online, we are coming to understand that when labels or narratives are decontextualized and amplified, we begin to internalize and adopt them, sometimes regardless of their accuracy or how savvy we believe ourselves to be.\textsuperscript{9} Moreover, what is blunt or vitriolic generally scales further and faster than what is nuanced or measured.\textsuperscript{10} As a label, fake news is arguably becoming so entrenched and normalized that it might ease the way for other terms that rhetorically marry the press to falsity, bias, and laziness—like “pink slime journalism”—to slip into our everyday discourse.\textsuperscript{11}

If protecting the press was the only goal of curbing anti-press rhetoric, that would be enough. But there is another reason to curb such rhetoric. How we talk about the press plays into how we tackle one of the biggest challenges of our networked age—stemming information pollution. Fundamental to this effort is separating accurate information from false, trusted sources from manipulated ones, and journalism from propaganda and marketing. If we use labels that conflate these categories, we make a daunting task harder.

As we barrel toward one of the American press’s biggest challenges of this century—reporting on the 2020 presidential election—we need to provide the press every possible support. Taking care in how we talk about the press should be part of that effort.

II. THE POWER OF NAMING

The belief that the act of naming brings with it great power is one that stretches across eras, religions, cultures, and academic disciplines.\textsuperscript{12} We could start with the Book of Genesis and how God’s naming of light with the


\textsuperscript{11} See \textit{infra} Part II.

\textsuperscript{12} See Loren Graham, \textit{The Power of Names: In Culture and in Mathematics}, 157 PROC. AM. PHIL. SOC’Y 229 (June 2013).
command—“Let there be light”—resulted in illumination.\textsuperscript{13} We could look at the fairy tale Rumpelstiltskin, in which an impish man lords over a miller’s daughter until she is able to learn the man’s name.\textsuperscript{14} We could look to the Russian-French mathematician Alexander Grothendieck, who, it was said, “had a flair for choosing striking evocative names for new concepts; indeed he saw the act of naming mathematical objects as an integral part of their discovery.”\textsuperscript{15} Evidence is rich that great power inures in the act of naming.

But naming goes beyond giving us the power to control. Names tend to shape how we understand things. As the Heisenberg principle posits, the very act of looking at something changes it. This phenomenon extends to linguistics as well, according to marketing scholar Adam Alter.\textsuperscript{16} “[A]s soon as you label a concept, you change how people perceive it,” Alter says.\textsuperscript{17}

This change in perception can be in service of humanity and freedom. For example, blogger and cultural critic Maria Popova has written that to name something is “to confer upon it the dignity of autonomy while at the same time affirming its belonging with the rest of the namable world; to transform its strangeness into familiarity, which is the root of empathy.”\textsuperscript{18} It can also be in service of beneficial progress. For example, naming is a precursor to our ability to problem-solve. Rebecca Solnit, a feminist writer and historian, has said, “When the subject is grim, I think of the act of naming as diagnosis. Though not all diagnosed diseases are curable, once you know what you’re facing, you’re far better equipped to know what you can do about it.”\textsuperscript{19}

But as naming can be generative, it can likewise be oppressive. Names can be used to minimize, defame, and distance or other. Law gives us a host of shameful—and current—examples of this. An entire section of the United

\textsuperscript{13} Id. at 229.
\textsuperscript{14} Id. at 231.
\textsuperscript{15} Id. (citing A. Jackson, \textit{Comme Appelé du Néant}, 29.5 \textit{NOTICES AM. MATHEMATICAL SOC’Y} 173–78 (1974)).
\textsuperscript{17} Id.
\textsuperscript{19} REBECCA SOLNIT, \textit{CALL THEM BY THEIR TRUE NAMES} 1 (2018); Maria Popova, \textit{Rebecca Solnit on Rewriting the World’s Broken Stories and the Paradigm-Shifting Power of Calling Things By Their True Names}, \textit{BRAIN PICKINGS} (Oct. 18, 2018), https://www.brainpickings.org/2018/10/18/rebecca-solnit-call-them-by-their-true-names/ [https://perma.cc/P5W9-RLGW].
States Code is titled “Aliens and Nationality.” And our gun laws refer to “mental defectives.”

As evidenced by these examples, damaging naming practices can have cultural staying power. According to Lucy Jewell, a scholar of law and rhetoric, “[H]armful rhetoric used to describe racial minorities and other subordinated groups produces toxic thought patterns that can become entrenched in the public mind.” Thus, to name, or to misname, something has tremendous significance. It shapes how we understand what is named, how we value it, and how we consider its possibilities. This is all true when it comes to the names we use to refer to the press.

III. CALLING THE PRESS NAMES

The story of how “fake news” became popular is a story of metastasizing meaning. According to Merriam-Webster, the term’s original meaning was a literal one. First used in the late nineteenth century, fake news was the sum of its parts. It meant false information published by the press.

Skip ahead more than a century, and the term reemerged. Perhaps reflecting the popularity of the spoof magazine The Onion, in the 1990s the term fake news meant news satire. Then, as information migrated online, the meaning of fake news again shifted. In 2014, Craig Silverman, a journalist documenting misinformation, discovered a false story describing a Texas town that had been quarantined after a family contracted Ebola. In an apparent attempt to make it look like a news article, the story included a made-up quote attributed to a hospital official. Silverman fired off a tweet linking to the false story and saying, “Fake news site National Report set off a measure of

21 See 18 U.S.C. § 922. This statute, in discussing who is barred from gun ownership, also happens to refer to an “alien” who is “illegally or unlawfully in the United States.” Id.
24 Id.
26 See id.
28 Id.
panic by publishing fake story about Ebola outbreak . . . Scumbags.”

With Donald J. Trump’s election as president, the path of fake news developed another well-known fork. About a week before his inauguration, Trump responded to a question from CNN’s Jim Acosta by saying, “You’re fake news.” Around the same time, the President adopted fake news as a Twitter mantra. At most recent count, the President had tweeted the terms “fake news,” “fakenews,” or “fake media” more than 700 times since his inauguration. And he has inspired copycats among other government officials, both in the United States and abroad, who also brandish the term as a sword. In 2017, use of the term fake news rose 365%. Beyond these modern uses of fake news (as satire, disinformation, and weapon), more permutations exist. One study examining academic articles that use the phrase fake news teased out six different meanings of the term.

Fake news is a jellyfish of a term—squishy and stinging. That sting is obvious when “fake news” is wielded as a weapon against the press. And

---

29 Id.
30 See id. In advance of the 2016 presidential election, “fake news” continued to be used to describe disinformation churned out by “hoax sites and hyperpartisan blogs.” See Craig Silverman, This Analysis Shows How Viral Fake Election News Outperformed Real News on Facebook, BUZZFEED NEWS (Nov. 16, 2016), https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook [https://perma.cc/U8QT-8ATU]. These sites produced false stories with such titles as “Pope Francis Shocks the World, Endorses Donald Trump for President” and “WikiLeaks CONFIRMS Hillary Sold Weapons to ISIS.” Id.
31 Wendling, supra note 9.
32 See id.
35 See Flood, supra note 6.
although it is necessary to call out those who are weaponizing the term, other people—especially journalists—have ably done this.\footnote{See, e.g., A.G. Sulzberger, The Growing Threat to Journalism Around the World, N.Y. TIMES (Sept. 23, 2019), https://www.nytimes.com/2019/09/23/opinion/press-freedom-arthur-sulzberger.html [https://perma.cc/AG3W-XVZ2] ("[W]hen the president decries ‘fake news,’ he’s not interested in actual mistakes. He’s trying to delegitimize real news, dismissing factual and fair reporting as politically motivated fabrications.").} Instead, my concern here is with those who likely do not intend harm when using the term. These individuals may even be avid supporters of the press. But regardless of intent, harm is still harm. As communications scholar Whitney Phillips has pointed out with respect to information pollution more generally, “The impact of industrial-scale polluters online—the bigots, abusers, and chaos agents, along with the social platforms that enable them—should not be minimized. But less obvious suspects can do just as much damage.”\footnote{Phillips, supra note 9.}

These less obvious suspects perpetrate harm in a variety of ways. One way is suggesting fake news is actually a type or subset of news. For example, the Wikipedia page for “fake news” begins by saying fake news is “a form of news.”\footnote{Fake News, WIKIPEDIA (last updated Apr. 20, 2020, 6:34 PM), https://en.wikipedia.org/wiki/Fake_news [https://perma.cc/R7RU-BQFJ].} Likewise, in a post on its website, Merriam-Webster indicates it is not planning to add fake news to its dictionary because fake news is “a self-explanatory compound noun” with “an easily understood meaning.”\footnote{The Real Story of “Fake News,” supra note 23.} According to Merriam-Webster, “Fake news is, quite simply, news (‘material reported in a newspaper or news periodical or on a newscast’) that is fake (‘false, counterfeit’).”\footnote{Id. at 1.} But defined this way, fake news is an oxymoron—and a damaging one.\footnote{Id. This report does have the words “Fake News” in its title, but those words are crossed out with a red slash on the cover page. Id. at 1.} As a United Nations report on disinformation explained, “[N]ews’ means verifiable information in the public interest, and information that does not meet these standards does not deserve the label of news.”\footnote{See Cherilyn Ireton & Julie Posetti, Journalism, “Fake News,” and Disinformation, UNESCO 7 (2018), https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf [https://perma.cc/65RY-KBWF].} That is, news is necessarily not fake. It may contain inaccuracies. It may lack context. But journalists are not intending to deceive. By suggesting journalists are taking us for a ride or the press is churning out false facts, this use of the term fake news linguistically links the press with falsity.\footnote{Put another way: “If it’s fake, it’s not news.” Adriaan Basson, If It’s Fake, It’s Not News, NEWS24 (July 6, 2017) https://www.news24.com/Columnists/AdriaanBasson/lets-stop-talking-about-fake-news-20170706 [https://perma.cc/QYU6-3EC6].}
Scholars and journalists may also unintentionally perpetrate harm. Although they have become more attuned to the hazards of using “fake news” (BuzzFeed’s Silverman says he now cringes every time he hears it45), some still invoke it in ways that feel gratuitous. For example, the title of a paper by Dartmouth researchers asks: “Real Solutions for Fake News?” Yet the paper then goes on to say that because the term fake news is “frequently used in imprecise and confusing ways,” the body of the paper will use other terminology.46 Setting aside the clearly good intentions of the authors, the title feels designed to grab attention. Likewise, a recent article in The Atlantic headlined The Conservatives Trying to Ditch Fake News is about an effort to create journalism for a conservative audience.47 Yet the body of the article does not use the term fake news at all.48

The continued popularity of “fake news” is not surprising. As a phrase, its two-syllable simplicity combined with its fuzzy meaning plays to the reflexivity and shallow thinking our frenetic online spaces encourage. Fake news can be slapped on all sorts of content, and it disparages on contact. This may be its appeal, but it is also its danger.

And the longer it enjoys popularity, the more entrenched the term fake news becomes. There may be no greater testament to this than its inclusion in a 2020 Super Bowl advertisement.49 The ad for Amazon’s Alexa includes an exchange in which a man asks a newsboy, “What’s today’s news?” and the newsboy responds, “Doesn’t matter. It’s all fake.”50 The ad’s use of fake news

45 Silverman, supra note 27.
48 Id. Another example is a recent article in Harvard’s Nieman Lab that uses “fake news” in the subheading but not at all in the article. See Mike Caufield, Ctrl-F: Helping Make Networks More Resilient Against Misinformation Can Be as Simple as Two Fingers, NIEMAN LAB (Jan. 29, 2020, 9:35 AM), https://www.niemanlab.org/2020/01/ctrl-f-helping-make-networks-more-resilient-against-misinformation-can-be-as-simple-as-two-fingers [https://perma.cc/VD83-YLDK]. The article’s subheading states: “Sometimes it’s the sort of basic Internet skill you might take for granted—like knowing how to search a web page—that can stop someone from sharing fake news.” Id.
49 See Amazon Super Bowl Commercial 2020 - #BeforeAlexa, YOUTUBE (Jan. 29, 2020) https://www.youtube.com/watch?v=RF9t2rFmTVE [https://perma.cc/B7EF-EK4K].
50 Id. It is hard to tell if the reference is made ironically, especially given that the newsboy shakes his head when the questioner cannot stop laughing at the “fake news” reference. See id. The lack of clarity, however, is part of the danger inherent in the term. Recognizing that danger, the director of communications for the Reporters Committee for Freedom of the Press tweeted her disappointment about the ad. Jenn Topper (@jenntopper), TWITTER (Feb. 2, 2020,
as a punch line is more concerning than funny. (It is especially concerning
given that Amazon’s owner, Jeff Bezos, also owns The Washington Post.\textsuperscript{51})

As described, naming practices shape how we think about people, ideas, and
institutions. Moreover, as those studying disinformation have confirmed,
when people are exposed to false information again and again, that
information begins to feel true—so true that it is believed even when people
are shown evidence of its falsity.\textsuperscript{52} Thus, even corrective efforts to
demonstrate that news is the product of journalism—a method aimed at
unearthing, contextualizing, and communicating truth—may be unsuccessful.

We should also be concerned that fake news may not be the last of its
ilk. It is possible this type of term—one linking journalism to falsity—will
proliferate. “Pink slime journalism” could be next. In December 2019, the
Tow Center for Digital Journalism at Columbia University published a report
entitled “Hundreds of ‘Pink Slime’ Local News Outlets Are Distributing
Algorithmic Stories and Conservative Talking Points.”\textsuperscript{53} The report, which
describes the mushrooming of 450 “partisan outlets masquerading as local
news organizations,” does more than the headline to separate “pink slime”
from “news.”\textsuperscript{54} But it is still possible to see how the phrase (or others like it)
could become the same type of weaponized, bloppy, value-laden smear as
“fake news.”

9:05 PM), https://twitter.com/jenntopper/status/1224151948812791809
[https://perma.cc/ER8F-4C8Q].
\textsuperscript{51} See Taylor Telford, Jeff Bezos Might Lose His Title As World’s Richest Person, WASH.
might-lose-his-title-worlds-richest-man/ [https://perma.cc/P5YE-X6QV].
\textsuperscript{52} Phillips, supra note 9.
\textsuperscript{53} Priyanjana Bengani, Hundreds of “Pink Slime” Local News Outlets Are Distributing
Algorithmic Stories and Conservative Talking Points, COLUM. JOURNALISM REV. (Dec. 18,
2019), https://www.cjr.org/tow_center_reports/hundreds-of-pink-slime-local-news-outlets-
are-distributing-algorithmic-stories-conservative-talking-points.php
[https://perma.cc/KX5R-ZJMW].
\textsuperscript{54} For further discussion of the term “pink slime journalism,” the report links to a blog post
about an episode of the radio show This American Life. That radio-show episode concerned a
company named Journatic that created “hyperlocal” journalism using automation as well as
employees based in the Philippines writing under false bylines. The same Journatic employee
featured in the This American Life episode told the journalism nonprofit Poynter, “I feel like
companies like Journatic are providing the public ‘pink slime’ journalism.” Id. (linking to Dan
Kennedy, Exposing the “‘Pink Slime’ Journalism” of Journatic, MEDIA NATION (July 5,
2012), https://dankennedy.net/2012/07/05/exposing-pink-slime-journalism/
[https://perma.cc/W8WK-EC4J]; Anna Tarkov, Journatic Worker Takes “This American
Life” Inside Outsourced Journalism, POYNTER (June 30, 2012), https://www.poynter.org/reporting-editing/2012/journatic-staffer-takes-this-american-life-
inside-outsourced-journalism/ [https://perma.cc/D8EQ-36SQ].
IV. WHY WHAT WE CALL THE PRESS MATTERS

Naming is of vital importance when it comes to the press. A great deal of press power springs not from the law but from language. Using—even without ill intent—language that has the potential to undermine the press is a risk.

For those who believe in a free press, there is solace in thinking the Constitution will protect the press. Journalists, for one, regularly invoke the First Amendment as a guardian. But although the press is named in the First Amendment, the Constitution’s power to protect the press is constrained. The First Amendment protects the press from Congress making a law infringing on press freedom, but it does not grant the press any affirmative rights. The Supreme Court has stated that the press has no special protections over and above those of any other speaker. And the First Amendment only protects the press from government overreach. It does not protect the press from private action.

Beyond the First Amendment, press power rests on what media-law scholars RonNell Andersen Jones and Sonja R. West have called other “pillars.” These include, among others, the press’s financial strength, the public’s trust in the press, and customs and norms, such as providing press access to events and public officials. These pillars are interdependent, and, in this moment, they are compromised. The press’s advertising-based business model is failing as technology platforms vacuum up advertising dollars.

---


56 See U.S. CONST. amend. I.

57 See id. (“Congress shall make no law … abridging the freedom of speech, or of the press”).

58 See Citizens United v. FCC, 558 U.S. 310, 390 n.6 (2010) (Scalia, J., concurring) (dismissing as “passing strange” the belief that the press should receive special constitutional protection); Branzburg v. Hayes, 408 U.S. 665, 704 (1972) (“Freedom of the press is a ‘fundamental personal right’ which ‘is not confined to newspapers and periodicals.’”).

59 See U.S. CONST. amend. I; Turner Broad. Sys. v. FCC, 512 U.S. 622, 685 (1994) (O’Connor, J., concurring in part and dissenting in part) (noting that “the First Amendment as we understand it today rests on the premise that it is government power, rather than private power, that is the main threat to free expression”).


61 See id.

Many press outlets are struggling to profit. The public’s trust in the press is not at an all-time low, but it is close. Only forty-one percent of Americans say they trust the media. Perhaps most alarmingly—because of its swiftness and speed—norms and customs are collapsing. The White House has discontinued press briefings. As of this writing in February 2020, it has not held a formal press briefing since March 2019. The Trump Administration has pulled press passes and otherwise denied journalists access to officials seemingly in retaliation for negative coverage. And the executive branch is not alone in demonstrating disdain for the press. During the Senate impeachment trial of President Trump, Capitol Police gave senators cards cueing them with language to use if the senators sought to avoid talking to journalists, including, “Please get out of my way” and “You are preventing me from doing my job.”

The term fake news and other anti-press rhetoric contribute to the pillars’ collapse. Repetition of the term is like unleashing groundhogs to burrow under the pillars of press freedom. The groundhogs may not singlehandedly bring them down, but they riddle the ground with holes, destabilizing it.

The effect of this destabilization may be greatest on the trust pillar. As researchers at the University of Texas concluded, “exposure to talk about fake news may lower individuals’ trust in media and lead them to identify real news with less accuracy.” Researchers primed subjects by showing them tweets

68 Emily Van Duyn & Jessica Collier, Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media, 22 MASS COMM’C’N & SOC’Y 1, 44 (2019); Daniel Funke,
referencing “fake news” before the subjects read articles. Even tweets that merely included article headlines with the words “fake news” tended to lessen trust.69

Beyond trust, anti-press rhetoric has also been blamed for creating an environment so rife with hate that it subjects journalists to harassment, threats, and even death. The Committee to Protect Journalists has blamed President Trump’s tweets for giving “cover to autocratic regimes” and notes it is aware of several U.S. journalists “who say they were harassed or threatened online after being singled out on Twitter by Trump.”70 And after the murder of five journalists while they worked in the Annapolis, Maryland, Capital Gazette newsroom, numerous journalists blamed pervasive anti-press rhetoric.71 This rhetoric includes “fake news,” but it also goes beyond it to include epithets such as “enemy of the people.”72

Because of their interdependence, the tumbling of one pillar could precipitate the tumbling of them all. Undermining the public’s trust in the press can hurt the press’s bottom line and, consequently, justify the erosion of norms and customs. As the New York Times editorial board wrote about the term fake news: “The capacity of news organizations to produce [hard-hitting] journalism—and to reach an audience that will listen—is contingent and


69 Duyn & Collier, supra note 68 at 35, 42; Funke, supra note 68. Notably, and perhaps seemingly at odds with the Duyn and Collier study, a 2019 study found that President Trump’s tweets about fake news may actually cause readers to believe the press is more credible. See Daniel J. Tamul et al., All the President’s Tweets: Effects of Exposure to Trump’s “Fake News” Accusations on Perceptions of Journalists, News Stories, and Issue Evaluation, MASS COMM’N & SOC’Y, 2019, at 7, 24. Yet, a synopsis of the study by Harvard’s Shorenstein Center indicated that the findings “cannot be generalized beyond the individuals who participated” in the studies—about 2,000 people, more than half of whom were undergraduate students. Denise-Marie Ordway, Fake News and Fact-Checking: 7 Studies You Should Know About, SHORENSTEIN CTR. ON MEDIA, POL., AND PUB. POL’Y (Jan. 13, 2020), https://journalistsresource.org/studies/society/news-media/fake-news-fact-checking-research-2019/ [https://perma.cc/SG72-JEN8].

70 Sugars, supra note 33.


Using the term fake news is one more contribution to the undermining of the press.

V. DIVORCING FAKE FROM NEWS

It is probably impractical and maybe even unhelpful to argue that the phrase fake news should never be used. For example, using the term in a way that does not unnecessarily highlight it and immediately provides context about its imprecision or using “fake news” to critique the term itself both seem legitimate and even beneficial. Yet, some self-restraint is in order. Before we mentally reach for fake news (or any other term that uses a broad brush to paint the press unfairly), we should engage in some strategic silence. In a journalistic context, strategic silence calls for consideration of the public good in deciding whether or not to share information, especially online.\footnote{See Joan Donovan & danah boyd, Stop the Presses? Moving from Strategic Silence to Strategic Amplification in a Networked Media Ecosystem, AM. BEHAV. SCIENTIST, Sept. 29, 2019, at 1 (defining strategic silence, in part, as the “use of editorial discretion for the public good”).}

In other words, we need to pause, mentally generate some of the friction largely absent in our online spaces, and consider whether a more precise term could substitute.


It includes terms like disinformation, misinformation, and malinformation.\footnote{Id.}

It does not include fake news, which Wardle, along with blogger and researcher Hossein Derakhshan, describe as “woefully inadequate to describe the complex phenomena of information pollution.”\footnote{Claire Wardle & Hossein Derakhshan, Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making, COUNCIL EUR. 5, 15 (Sept. 27, 2017), https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c [https://perma.cc/Q2Q6-7GYK].}

The substitutions may not be as catchy, but they are also not as poisonous. For example, in a Twitter thread, Renee DiResta, the technical research manager at Stanford Internet Observatory, referred to “‘pink slime’ content farms that look like journalism.”\footnote{Renee DiResta (@noUpside), TWITTER (Feb. 8, 2020, 11:29 AM), https://twitter.com/noUpside/status/1226181774390087681 [https://perma.cc/5PQL-CBZM].} That description is eight words to
pink slime journalism’s three, but it also helps make the nuanced point that this content is not news or journalism even if it is intended to look like it is.

Shunning anti-press rhetoric, especially fake news, is also a component of a broader effort to remedy information pollution. Along with “real-world” pollution and climate change, the problem of information pollution is rapidly evolving into one of the greatest humans face—impacting our mental and physical health, our elections, and our democracy.\(^79\) We need to be able to name and define the components of this issue precisely. We also need to allow for those names and definitions to evolve as the underlying challenges morph and, hopefully, our grasp of them simultaneously tightens.\(^80\) If, as Rebecca Solnit said, naming is an act of diagnosis, the term fake news is not only obscuring the disease but is also feeding it.\(^81\)

Another benefit of sweeping away anti-press rhetoric is the creation of space for press-affirming rhetoric. Journalism scholar Nikki Usher has argued that in rejecting the term fake news there exists “an interesting branding opportunity to possibly restore trust in journalism.”\(^82\) Journalists are engaging in this to some degree. As one example, Usher pointed to a *New York Times* campaign of handing out “truth buttons” at events—pins saying things like “The truth is hard” and “Truth: It’s more important now than ever.”\(^83\) Far beyond this, journalists are working in numerous ways to rebuild trust and faith in their discipline. For example, researchers and academics at the University of Texas’s Center for Media Engagement are testing the impact of what they call a “Behind the Story card” offering “information about why and how a story was written.”\(^84\) The American Press Institute and the Reynolds Journalism Institute also have a project, Trusting News, that provides


\(^80\) See Wendling, supra note 9 (quoting Claire Wardle as saying, “If we're going to start thinking of ways we can intervene, we're going to have to have clear definitions”).

\(^81\) See Solnit, supra note 19.


Journalists trainings and resources regarding how to build more trust in their reporting and stories.\textsuperscript{85} Other efforts abound.\textsuperscript{86}

But individual press advocates must also play a role. The press needs those who believe in its work to amplify and generate these pro-press messages, whether through conversation, social media, public speaking, or scholarly articles. As Whitney Phillips counsels with respect to remedying information disorder, we need to understand our own agency. Phillips advocates that “[t]o have any hope for a different future, we must survey the landscape, consider where our own bodies stand, and ask: How might what I do here affect what happens over there?”\textsuperscript{87} Although various means exist for supporting the press (including financial ones), we can each support high-quality journalism and the press with our words.

VI. CONCLUSION

“It is important to remember that language itself is a moral medium,” wrote philosopher and novelist Iris Murdoch.\textsuperscript{88} “[A]lmost all uses of language convey value.”\textsuperscript{89} As a phrase, fake news devalues. When used uncritically—without explanation or interrogation—it has the potential to undermine an already besieged press.

It is, of course, right to critique the press. The press, the journalists who comprise it, and the journalism they engage in are all imperfect. But fake news is not reasoned or thoughtful critique. It is a hazy and often hastily-applied label that can erode trust. Harm can result regardless of whether the speaker intended harm.

The press is already unthinkably fragile. Law may provide a thin layer of bubble wrap around it, but it does not guarantee the press’s safe passage into the future. To protect the press, we must take care with the language we use to talk about it. That language can promote trust in and respect for the

\begin{footnotesize}
\begin{itemize}
    \item \textsuperscript{85} Trusting News Project: Helping Journalists Earn News Consumers’ Trust, TRUSTING NEWS, https://trustingnews.org [https://perma.cc/NUZ6-MFAU].
    \item \textsuperscript{87} Phillips, supra note 9 (emphasis in original).
    \item \textsuperscript{88} Maria Popova, Iris Murdoch on Storytelling, Why Art is Essential for Democracy, and the Key to Good Writing, BRAIN PICKINGS (July 18, 2018), https://www.brainpickings.org/2018/07/18/iris-murdoch-existentialists-mystics-philosophy-literature-art/ [https://perma.cc/4L38-ML4E].
    \item \textsuperscript{89} Id.
\end{itemize}
\end{footnotesize}
press and journalism as a method. Or it can suggest news is just another form of pollution in our damaged information ecosystem.

The right language is not only essential for the purpose of protecting the press so it can continue its democracy-promoting work. It also helps us to better understand the pollutants in our information environment and work to remediate them. And such language helps allow the press to be a force in that remediation, providing truthful, contextual, and newsworthy information to the public.
MAKING UP POLITICAL PEOPLE:
HOW SOCIAL MEDIA CREATE THE IDEALS,
DEFINITIONS, AND PROBABILITIES OF
POLITICAL SPEECH

Mike Ananny*

CITE AS: 4 GEO. L. TECH. REV. 351 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 351
II. FACEBOOK’S FACT-CHECKING PARTNERSHIP AS INFRASTRUCTURE... 354
III. THREE WAYS PLATFORMS SEE PUBLIC LIFE AS/THROUGH
     COMMUNICATION INFRASTRUCTURE ........................................... 357
     A. Information Ideals of Citizenship ............................................ 357
     B. The Categories of Politics and the Politics of Categories ......... 359
     C. Governing Scale through Probability ..................................... 362
IV. CONCLUSION ........................................................................... 365

I. INTRODUCTION

Especially in the wake of the 2016 U.S. presidential election, there is increasing debate about how and why to regulate political speech posted on social media platforms, and whether such regulation is possible or desirable. Initially, such debates focused on the concept, production, and circulation of “fake news.” As politicians like President Trump co-opted the term to attack

---

* Associate Professor at the University of Southern California’s Annenberg School for Communication and Journalism; Ph.D., Stanford University, 2011. Thanks to: Prof Julie Cohen for the invitation to participate in and develop this article for Georgetown Law’s conference “Election Integrity in the Networked Information Era;” Professors Erin Carroll and Julie Cohen for their thoughtful feedback; and the staff at the Georgetown Law Technology Review for their patient and helpful editorial guidance.

journalists and insulate themselves from criticism, and as academics argued against the term’s imprecision, historical complexity, and political weaponization, this term fell out of favor. In lieu of discussing “fake news,” scholars and public commentators gradually began talking about the role that “misinformation,” “disinformation,” or “computational propaganda” play in public discourse.

Rarer, though, in debates about how to name and trace this problematic class of speech are more fundamental conversations about what type of public life is presumed or desired. More precisely: What types of public life does such a focus on truth or falsity take for granted, and what type of public life is created by focusing on the truth or falsity of information? While many scholars assume the existence of “fake news” as a powerful political phenomenon and focus on its spread, effects, and defenses, others have questioned the “myth of the attentive public” that fake news ostensibly harms, reminded us of persuasion campaigns’ minimal effects, and questioned why scholars seem to be focusing on questions of information quality and circulation instead of longstanding political issues like race, class, and identity.

In this Paper, I want to depart slightly from scholarship that frames debates about contemporary, online political speech in terms of information, persuasion, effects, or even identity and instead ask: What assumptions do platform infrastructures make about public life, and what alternative forms of public life to do those assumptions foreclose? I focus on these infrastructures not because I see them as simple channels for delivering content that sways minds or elections, but because they are simultaneously instrumental and symbolic. Yes, they expose audiences to content, but they also serve as powerful sociotechnical imaginaries: “Collectively held, institutionally

---


6 See, e.g., Computational Propaganda (Samuel C. Woolley & Philip N. Howard eds., 2018).


9 See DANIEL KREISS, Prototype Politics 93–100 (2016).
stabilized, and publicly performed visions of desirable futures” enacted through often invisible technologies.10

Science and Technology Scholars often focus on the public significance and power of built infrastructures—information systems and social arrangements that create the conditions under which collective meanings and public life are made.11 In the same tradition, I posit that today’s dominant view of online political speech—and the dominant critiques thereof—rests on three largely unquestioned assumptions. Firstly, citizens are *information processors* and publics are *information products*. Secondly, public life rests on *taken-for-granted categories* that social media platforms have a vested interest in creating, naturalizing, and policing. Lastly, much of political speech is governed by largely invisible *logics of probability* that platforms use to justify a scale of data that their business models require.

To develop this argument and ground this critique, I focus on a partnership that Facebook formed with five U.S. news and fact-checking organizations in the wake of the 2016 U.S. presidential election. Drawing on interviews and fieldwork with key partnership participants, I show how this partnership was a form of political speech infrastructure that assumes an “information ideal” of citizenship, that rests upon proprietary and computed categories of political speech, and that uses probability as a tool to govern speech and as a rhetorical defense against the power of Facebook’s scale.

Although a complete discussion of the concept is beyond the scope of this Paper, I use the term “infrastructure” in the tradition of Science and Technology Scholars who largely define it as an often invisible set of relationships among people and materials (e.g., algorithms and databases) that create the conditions under which sociotechnical systems are seen to be working or failing.12 In addition to the more visible and formalized rules that govern organizations’ and people’s use of technologies, these relationships create the tacit norms, taken-for-granted practices, collective rhythms, shared terminology, and agreed-upon roles that make complex systems—from transport systems and power plants to advertising algorithms and sensor networks—seem stable, predictable, governable, and even boring. When such infrastructures are “inverted”13—when the usually invisible *relations* among their human and nonhuman actors are traced and analyzed—it becomes possible to see how information systems are actually precarious, ongoing

---

achievements that require a great deal of negotiation, coordination, and compromise. Infrastructures are anything but natural. They could be driven by different relationships, different definitions, and different tradeoffs. At any given moment, an infrastructure’s form and function emerge from subtle and complex negotiations among people and machines, many of whom are not consciously aware of or reflecting upon the types of power, language, expertise, and values that are keeping their infrastructures alive. In light of Latour’s claim that “technology is society made durable,”14 the concept of infrastructure becomes a way of (a) seeing technology as intertwined, mutually defining relationships between people and materials, and (b) asking which other societies might be possible, if only we had different relationships and different technologies.

Increasingly, scholars are using such an infrastructural lens to understand information work—everything from elections,15 earthquakes,16 canals,17 and space exploration,18 to press freedom,19 artificial intelligence,20 television distribution,21 and fact-checking.22 This way of seeing technologies focuses little on the “effects” that media have on people. Instead, such an approach cares more about what kind of social, cultural, and public worlds are created though intertwined people and objects. In this vein, I not only claim that Facebook’s partnership with news organizations can best be seen as infrastructural, I also want to read this infrastructure for the assumptions it makes about what public life is and should be.

II. FACEBOOK’S FACT-CHECKING PARTNERSHIP AS INFRASTRUCTURE

I base the analysis in this claim and this reading on a case study I conducted of Facebook’s partnership with five U.S. news and fact-checking

15 See generally Kreiss, supra note 9.
16 See generally Megan Finn, DOCUMENTING AFTERMATH (2018).
18 See generally Janet Vertesi, Seamful Spaces: Heterogeneous Infrastructures in Interaction, 29 SCI. TECH. & HUM. VALUES 264 (2014).
19 See generally Mike Ananny, NETWORKED PRESS FREEDOM (2018).
21 See generally Joshua A. Braun, THIS PROGRAM IS BROUGHT TO YOU BY (2015).
organizations. The partnership has since received significant journalistic attention, has changed, and is, in many ways, no longer the same network that I studied in 2017–2018. The point of this Paper, though, is neither to give a contemporaneous account of the network as it exists today, nor to evaluate the partnership’s success or its impact on the spread of misinformation. Rather, my aim is to use the network as an object to think with about how a major social media platform understands political speech, sees its public purpose, and configures its infrastructures in ways that make some types of public life more or less possible.

My 2018 report gives fuller context on the partnership’s origins and dynamics, but it is sufficient here to say that the collaboration was formed in the immediate aftermath of the 2016 U.S. presidential election, spurred on by a combination of factors. Whether the effect of its misinformation on the election was real or imagined, Facebook was caught in a major public relations scandal as people inside and outside of the company called for some kind of change to how the platform circulated content. At about the same time, the International Fact-Checking Network (IFCN)—a unit of the Poynter Institute that convenes fact-checkers worldwide—wrote an open letter calling for Facebook to provide more transparency and accountability in how it circulated misinformation. Facebook and five U.S. news and fact-checking organizations that were signatories to the IFCN’s code of principles formed a partnership in which fact-checkers could “identify a story [that appeared on Facebook] as fake,” flag that story as “disputed,” and provide a “link to the corresponding article explaining why.” Stories that fact-checkers flagged as disputed could then potentially “appear lower in News Feed.” Fact-checkers were hungry for access to the raw materials that they needed for their work—misinformation—and one of the principal suppliers of those materials—Facebook—was offering them a type of insider access.

This workflow was structured around a dashboard that Facebook built and then let its fact-checking partners access. Fact-checkers loaded the dashboard and reviewed the list of stories that Facebook had posted to it (based

---

24 See id.
25 Id.
27 See Ananny, supra note 23.
28 Although I was never allowed to see the dashboard, the description of it here was confirmed by multiple partners. See id.
on “reports from [their] community, along with other signals”). The dashboard organized the stories in various ways, including by date, URL, and a “popularity” score that fact-checkers said Facebook never fully explained to them. Fact-checkers would choose a story to investigate, write their report, and then append their report to that story’s dashboard entry. From there, Facebook would analyze the stories, the fact-check report, and other undisclosed signals. Some news organizations had a dedicated person responsible for choosing stories to fact-check off the Facebook dashboard, while other news organizations shared the role among multiple people. All partners stressed that they treated this content no differently than other stories they fact-checked. In a leaked email obtained by BuzzFeed News—later confirmed by Facebook—the company claimed that “once we receive a false rating from one of our fact checking partners, we are able to reduce future impressions [of that story] on Facebook by 80 percent.”

The partnership suffered from several challenges. For example, fact-checkers debated amongst themselves whether and how much money to accept from Facebook for participating in the collaboration. Additionally, the dashboard and workflow were routinely critiqued as opaque and overwhelming; the dashboard listed thousands of stories, and fact-checkers felt that their methodical practices were too slow for the deluge. Fact-checkers further complained that they knew too little about the impact of their work, that there was ambiguity regarding why certain sources but not others appeared on the dashboard, and that Facebook did not communicate with them often or clearly enough.

Although the partners worked to ameliorate several of these issues—and some partners left while others were added—it was clear that the partnership was infrastructural. It entailed norms, practices, and values among a variety of actors. Automated machine processes served up disputed stories, pushed the fact-checks back into an unseen system, and then reduced the circulation of those stories. The dashboard organized work and spurred debates and, unless you were the Facebook engineer who built the dashboard or the fact-checker who used it, much of the partnership was beyond your

29 See Ananny, supra note 23.
32 Ananny, supra note 23.
33 Id.
34 Id.
view, control, or even knowledge. Despite the attention it periodically received in the mainstream media and Facebook’s public touting of its success in curbing misinformation, the partnership was largely invisible to most people. Just as most train passengers think little about railroad ties and few Netflix watchers think about the broadband networks enabling their binges, the fact-checking partnership became part of the largely invisible speech infrastructure that heavily influences what gets said, who hears it, and whether it is believed, shared, and acted upon.

III. THREE WAYS PLATFORMS SEE PUBLIC LIFE AS/THROUGH COMMUNICATION INFRASTRUCTURE

A. Information Ideals of Citizenship

The infrastructure of Facebook’s fact-checking partnership had many rich dynamics. In particular, this Paper will focus on three things that the partnership—and platforms more generally—teaches us: How political communication technologies (1) see people, (2) defend categories, and (3) assume scales. To construct a normative critique of contemporary political communication—to ask if society is creating the kinds of publics that it wants and needs—one must develop precise interrogations of the infrastructures that structure political speech.

The fact-checking partnership was driven by an image of the ideal political person that mirrors what media historian Michael Schudson calls the “good citizen.”35 Emerging out of late nineteenth- and early twentieth-century progressive moments, this was a “citizenship of intelligence rather than passionate intensity”36 that foregrounded individual literacy and disciplined self-education—“a shift from sentiment to interest as the basis of politics.”37 This version of the good citizen is echoed in progressive models of journalism and deliberative theories of democracy that aim to give citizens the “objective” information that they need to vote, hold public officials accountable, and rationally debate policy alternatives.38 Indeed, fact-checking cultures often play into this information ideal by seeing themselves as verifiers of truths that citizens can use in whatever ways they like.39

36 Id.
37 Id. at 185.
38 See, e.g., MICHAEL SCHUDSON, WHY DEMOCRACIES NEED AN UNLOVABLE PRESS 12 (2008).
I heard exactly this ideal of the citizen in how partners talked about the partnership and their work. Fact-checkers spoke of their need to get at the massive amounts of misinformation that Facebook indexed, their investment in “truthful reporting,” and their desire to “reduce deception in politics.” Facebook said that it “cannot become arbiters of truth ourselves,” but could instead use the partnership as a way to “help people decide for themselves what to trust and what to share.”40 Fact-checkers and Facebook saw themselves as information brokers whose job was to increase the quality and validity of information circulating online.41

With only a small amount of dissent, fact-checkers and Facebook agreed that the misinformation its partnership identified should not be banned entirely from the platform (or the broader Internet).42 Some suggested that misinformation should be moved to an area separate from Facebook’s main newsfeed, while others stressed that if misinformation was presented alongside fact-checks then people could decide for themselves what to make of the stories. They all largely adopted a marketplace model of political speech. One fact-checker said that that people’s “motivation for sharing might be complicated . . . and [we] don’t want to police the Internet.”43 Another argued that as long as fact-checks accompanied misinformation, people would learn what was true, and as “a First Amendment supporter I find that cool.”44

The partnership’s dominant image was of “good citizens” critically informing themselves, publics rationally debating quality information, and speech marketplaces surfacing truths. This image of the literate, rational, information-exchanging citizen, though, aligns poorly with (a) how effectively emotional political content drives social media popularity45 and (b) the prominence of the “popularity” metric in the partnership’s own dashboard. Platforms thrive on emotional content, so it is unsurprising that “popularity” would be one of the dashboard’s keywords.

It could be that partners thought that their fact-checking could counter the popularity of emotional political content—i.e., they knew that citizens are

---

40 Mosseri, supra note 26.
41 See Ananny, supra note 23.
42 Id.
43 Id.
44 Id.
not rational information processors, but they wished they were and were trying to make them such by debunking popular misinformation. This wish contradicts a larger point made by scholars of evocative media, affective publics, and agonistic politics. Political emotion is immutable and sometimes desirable. Instead of trying to curb emotional content and make people into something they are not—a fool’s errand given social media’s love of affective engagement—partnerships and infrastructures like Facebook’s might acknowledge the immutability of emotion and instead ask how affective media could create public life that is defensible and desirable.

B. The Categories of Politics and the Politics of Categories

The information ideal of citizenship is not the only contingent and contentious concept underpinning the partnership and social media more generally. Facebook and other platforms rely heavily on several largely opaque and unexplored definitions and categories that are made to seem like natural and essential elements of online political speech.

The most central term is, in fact, a collection of terms related to falsity. In a single 2018 policy statement, Facebook product manager Tessa Lyons uses several seemingly interchangeable terms, including “false news,” “misinformation,” “fake stories,” “inauthentic content,” and “misleading content.” There is little clarity or precision in Facebook’s language.

To be sure, the phenomenon is by no means obvious or incontrovertible. Lucas Graves, one of the leading scholars of journalistic fact-checking, describes how messy and interpretive the practice can be. He finds that fact-checking is anything but a quick verification of pre-existing claims. Instead, it is a form of knowledge production in which “simple and settled questions seem to become more complicated and less settled on closer inspection,” as fact-checkers weigh different types of authority and expertise, consider incomplete and uncertain information, and interpret the assumptions and expectations of different audiences. Rarely do fact-checkers encounter the kind of neat ontological categories that platforms and machine learning algorithms assume. Facebook has a vested interest in outsourcing uncertainty to fact-checkers, whose judgments they can use to “improve our technology

---

49 Lyons, supra note 30.
50 See id.
51 Graves, supra note 39, at 69.
52 Id.
so we can identify more potential false news faster in the future.”

Though self-reflective fact-checkers know that their determinations can be contingent and contextual, Facebook requires “true” or “false” to be stable categories policed by professionals whom the platform can lean on for the legitimacy it needs to claim that its circulation of political speech is neutral and objective.

Fact-checkers balked, though, at one of the dashboard’s central categories: “Popularity.” As one fact-checker said, “We’ve asked [Facebook] a hundred ways to Sunday what ‘popularity’ means. We don’t know the mechanism they use to determine popularity.” Several fact-checkers questioned why they never saw widely popular mainstream news or conservative conspiracy sites like InfoWars listed on the dashboard. Fact-checkers also wondered whether they were missing important forms of misinformation that did not fit with Facebook’s calculation of “popular,” and speculated (without evidence) that Facebook might not define as “popular” stories that earned the platform a great deal of advertising revenue. Partly in response to partners’ concerns, Facebook added an “impact” metric to the dashboard, though fact-checkers similarly did not understand how this term was defined.

But even though fact-checkers questioned these terms and calculations, they seemed simultaneously grateful for them. They knew that popularity was an important phenomenon and seemed willing to let Facebook own the operationalizations of these concepts, or simply accepted that they could never persuade Facebook that such calculations were proprietary trade secrets that could not be discussed in detail. One fact-checker expressed gratitude to Facebook for “doing that work, providing views on what’s popular, what’s circulating.” And another acknowledged that even imperfect popularity metrics helped them allocate their time: “You don’t want to write about something that hasn’t gone viral because you don’t want to elevate its visibility. But if there’s something that is being widely circulated, then you want to debunk it.”

These debates about “true” versus “false,” “inauthentic” versus “verified,” and the meaning of “popular” are only a small part of Facebook’s larger assumptions about the stability and incontrovertible aspects of political speech. For example, in his 2019 speech outlining the platform’s work to

---

53 Lyons, supra note 30.
54 Ananny, supra note 23.
55 Id.
56 Id.
57 Id.
58 Id.
59 Id.
60 Id.
“prevent outside interference in elections and Facebook’s attitude towards political speech,” Facebook Vice President of Global Affairs and Communications (and former Deputy Prime Minister of the UK) Nick Clegg foregrounded—but left undefined—an almost endless number of key terms.61 Those terms included “politician” (does this include officeholders and office-seekers at all levels of government?), “newsworthy” (it depends upon the content’s “public interest value” and whether a region has a “free press”), “election” (it is unclear when Facebook thinks these start and stop), and “deepfake” videos (versus shallowly faked forms of other media that are presumably less newsworthy or of lower public value).62

To be sure, these terms are unstable and not easy to define. They shift depending on the context and the different values and assumptions of the speaker, and they are the subject of endless scholarly debate. My concern is not that Facebook has defined these terms incorrectly—people can and should disagree about them—but rather that Facebook has created operationalizations of these terms through largely opaque and poorly understood processes. Further, the inaccessibility these terms places them not only beyond critique—they become baked into the infrastructures of online political speech—but they can become naturalized in a way that gives them an entirely new type of political power. To function at the scale that its business model requires, Facebook needs to conceal its definitional work—to be sure, the platform’s public power and opacity is a problem—but my deeper concern is that Facebook is creating a seemingly natural set of terms and definitions that will become dominant categories of political speech and, thus, the categories used for collective self-governance. It is one thing not to be able to see inside the machinery that operationalizes and governs political speech, but it is quite another not to be able to debate the meaning of words like politician, newsworthy, public interest, or free press—or even know what words and categories make the platform work.

Facebook is not alone in making these categorical assumptions. News organizations often breezily switch between calling their audiences “users” versus “readers”; Twitter describes its “trending topics” as a neutral image of people’s interests, Google maintains a private whitelist of websites that it judges worthy of indexing on Google News, and journalists themselves often equate tweets and search-engine queries with “public opinion.”63

---


62 See id.

politics shrinks when we fail to acknowledge the politics of political categories.

This debate is not new. In her foundational paper “Do Categories Have Politics?” Science and Technology Scholar Lucy Suchman identified a class of “coordination technologies” that “provide canonical frameworks for the representation and control of everyday communication practices.”64 Her concern was that such technologies were based on a set of uncritiqued and largely invisible models that aimed to “discipline and control” people’s actions.65 After Suchman used Terry Winograd’s system to illustrate her argument, Winograd, doctoral advisor to Google co-founder Larry Page and later a company advisor, responded by largely dismissing her core critique as “sociopolitical drama,”66 but accepted that no system can “fully capture the richness of mental life or social interaction.”67 Standardization and coordinated action were essential if distributed activities were to accomplish a goal.68 Winograd’s solution to Suchman’s charge was to invite people to participate in the creation of computational systems, arguing that “design succeeds when it is grounded in the context and experience of those who live in the situation.”69

The problem with Winograd’s response to Suchman’s astute critique is that those of us who “live in the situation” that Facebook creates—who are “grounded in the context and experience” of the political speech the platform governs—often cannot see, understand, consent to, or counter its categories and definitions. The deep challenge is not only accessing these terms and holding those who dominate the political life that they describe accountable, but fighting against their privatization and naturalization.

C. Governing Scale through Probability

Touting the success of the fact-checking partnership, Lyons, the Facebook product manager, said that when partners “rate something as false, we rank those stories significantly lower in News Feed. On average, this cuts future views by more than 80%.”70 Several fact-checkers said that they were unsure what to make of this statistic, replying, “I don’t know how that number

64 Lucy Suchman, Do Categories Have Politics? The Language Action Perspective Reconsidered, 2 COMPUTER SUPPORTED COOPERATIVE WORK 177, 178 (1994).
65 Id.
67 Id. at 192.
68 Id.
69 Id. at 195.
70 Lyons, supra note 30.
is calculated,” “We have no public proof of that,” and, ironically, “I can’t fact-check that claim, and that’s a problem.” 71

Regardless of whether the statistic is true, it points to a larger dimension of how social media understands the regulation of political speech. Instead of banning misinformation or deleting problematic content, Facebook and other social media platforms rely instead on a probabilistic model of speech regulation. They never promise that users will not see content, they just say that it will be less likely. This rationale seemed to frustrate fact-checkers. It points to a new, probabilistic way to treat fact-checkers’ work that they did not fully understand or agree with. “We don’t get any information about these stories or what’s happened to them [after we fact-check them]. What they’re telling the public is no more than what we’re getting. We don’t know and we don’t understand how they’re using it.” 72

This probabilistic gray area was new because it was unlike how fact-checkers saw the value of their work and a different kind of debate than the one they were accustomed to. Fact-checkers told me that they were used to a debate about whether false information should be banned from a platform entirely (de-indexed and entirely removed as best the platform could) or whether false information should continue to appear but displayed alongside the fact-check. Some fact-checkers favored outright bans, others suggested that false information be placed in a separate area, and still, others were comfortable with their fact-checks appearing alongside false information. 73 But no one knew what to do with Facebook’s regulation-through-probability. Why was the 80% number celebrated? How were the remaining 20% of views distributed? Who was still seeing the content, and was it better or worse that they saw it more or less often than others? In his speech, Clegg said that Facebook would “demote” (not ban) “previously debunked content” that “politicians” shared, 74 so was the Facebook the 80% statistic the same number that would drive those demotions? Were there different percentages for politicians and non-politicians (however each is defined)? Did 80% represent a victory that satisfied Facebook, a target that their algorithms were trying to hit, or was it trying to move to a different number—85%? 90%? 97%? Presumably, the target is not 100%, because that would be a ban, and Facebook’s then-vice president for its News Feed product said “that we cannot become arbiters of truth ourselves.” 75 To Facebook, probabilities are governors, but not arbiters. They seem to see themselves as creating statistical

71 Ananny, supra note 23.
72 Id.
73 Id.
74 Clegg, supra note 61.
75 Mosseri, supra note 26.
likelihoods that exist below some threshold of chance that does not act as a deterministic judgment.

As with the discussion of political categories, the point here is not to advocate for an ideal number or to say that Facebook is succeeding or failing with this statistic. Rather, what emerged through the partnership was the idea that speech should be governed probabilistically, that fact-checkers’ work would somehow feed into those likelihoods in ways that they did not understand and could not verify, and that this type of statistical, algorithmic governance was less odious to the circulation of political speech than some other system would be.

I elsewhere discuss how this type of probabilistic governance is not unique to Facebook and suggest that it is a new and understudied logic common to many online platforms. For example, in 2019, YouTube said it would “begin reducing recommendations of borderline content and content that could misinform users in harmful ways.” Twitter’s Del Harvey said that, “Given the context of the scale we’re dealing with, if you’re talking about a billion tweets, and everything goes perfectly right 99.999% of the time, then you’re still talking about 10,000 tweets where everything might not have gone right.” And, after the ACLU discovered that Amazon’s Rekognition facial recognition system falsely identified twenty-eight members of the U.S. Congress as felons, the company said that it recommended that developers of law-enforcement applications set its API match metric at “a threshold of at least 95% or higher,” instead of the API’s default 80% confidence level.

Platforms’ business models need a scale that is ungovernable by anything other than statistical techniques. Although platform antitrust initiatives continue apace, it is inconceivable that all of the large-scale social media platforms will be broken up and become small enough for exclusively human curation and moderation. Given that scale is here to stay, it seems imperative that governance mechanisms like probabilistic speech regulation

---


receive greater critical attention. How are such thresholds, probabilities, and confidence intervals set? Are they explicit, humanly set limits, or are the variable outputs of machine learning systems designed to find different kinds of patterns? What happens when categories and probabilities intersect—e.g., when the likelihood of a video being a “deepfake” needs to be compared against the chance that bots are circulating a video, which has been shared with by someone who behaves “like” a politician, in a country that “usually” enjoys “press freedom?” As we try to protect the circulation of “political speech” during “elections,” what thresholds and confidences do we need to meet for people to see those races as legitimate? The complexities of probabilistic governance quickly spiral beyond comprehension and control. Is it possible to understand probabilistic systems well enough to slow their spirals?

IV. CONCLUSION

Politics often involves fighting about ideals, categories, and impacts. The type of politics unfolding in online environments is no different, except the ideals, categories, and impacts of contemporary, networked politics are playing out in infrastructures. Private, self-regulating corporations largely control these infrastructures. Even platform employees with full access find it difficult to trace and understand them. And scholars are only beginning to develop the techniques needed to critique the precise ways that speech infrastructures make some speech, and some forms of public life, more likely and desirable than others.

This all matters because, although there is a great deal of smart conversation about how and why to regulate platforms, there does not seem to be as much deep thinking about what kind of public life platform regulations aim to create. Is the ideal public a rational, deliberative, truth-seeking one that Facebook and its fact-checkers seem to want? Is it a participatory one that is less concerned about truth and more concerned about the exchange of opinions? Is it an aggregation of responses to polls, surveys, and questionnaires? Is it an agonistic one that cultivates disagreement, and manages shared consequences, without ever thinking that anything like consensus is ever possible? In all likelihood, it is some combination of these, but it is well past time to create platform regulations that move beyond an almost exclusive focus on marketplace models of speech and deliberative ideals of the public, to messier and normatively complex images of the public interest.80

One of the places that these images of public life appear is in the media infrastructures that govern online speech. These infrastructures are messy,
hidden, and seemingly boring things like content dashboards, structured data formats, political relationship management systems, and news advertising networks. Though journalism scholars are beginning to interrogate the epistemological dynamics of these infrastructures—how they make and circulate knowledge claims—there is still little work on these infrastructures’ normative power—their ability to “make up people” and the rules of politics and thus create public life.

I have tried to show here that there are (at least) three ways of making up political people: By implicitly adhering to an information ideal of the citizen and a marketplace model of speech, by letting platforms create and naturalize the terms and categories that political life depends upon, and by uncritically accepting the probabilistic governance systems that the business models of large-scale social media platforms seem to need.

To imagine different kinds of publics and sociotechnical futures—e.g., to see press freedom differently or create public media systems—we need to reject exclusively information-based ideals of political personhood, denaturalize social media categories, and trace, resist, and perhaps ban probabilistic governance that is beyond human oversight.

---

82 See Graves & Anderson, supra note 22, at 347.
86 See generally ANANNY, supra note 19.
USER CORRECTION AS A TOOL IN THE BATTLE AGAINST SOCIAL MEDIA MISINFORMATION

Leticia Bode*

CITE AS: 4 GEO. L. TECH. REV. 367 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ........................................................................................................................................ 367
II. PROBLEMS WITH MACHINE LEARNING-BASED CONTENT MODERATION .............................................. 370
   A. Empirical Problems with Automated Content Moderation on Social Media ............................................. 370
   B. Ethical Problems with Automated Content Moderation on Social Media ................................................. 372
III. TOWARDS A MISINFORMATION SOLUTION ......................................................................................... 374
IV. CONCLUSION ........................................................................................................................................ 378

I. INTRODUCTION

Misinformation is not a new problem. As long as information is valuable in helping people make decisions, there will be an incentive for third parties to manipulate that information in a way beneficial to their interests. The result is misinformation, which has taken many forms over the years.

Misinformation has a variety of definitions. In this Paper, misinformation is used to mean information that is “considered incorrect based on the best available evidence from relevant experts at the time.”¹ This definition does not discriminate based on intent—information is incorrect,

* Provost’s Distinguished Associate Professor in the Communication, Culture, and Technology program at Georgetown University. Dr. Bode’s research focuses on political and health communication, particularly in terms of how technology affects whether people are exposed to information, how accurate it is, and what they do with that information. Ph.D. University of Wisconsin—Madison, 2012. The author would like to thank Emily Vraga for years of thinking through many of these ideas.

regardless of whether its intent is to mislead\(^2\) (disinformation) or not (misinformation).\(^3\)

Although misinformation itself is not new, the form it has taken in recent years has changed. Technological innovation, broadly speaking, and social media, more specifically, allows for the rapid dissemination of both information and misinformation, and there is at least some evidence that the latter spreads more easily and more quickly than the former.\(^4\) Although the nature of the problem may be overstated—a recent study found that only 8.5% of people shared misinformation with their friends on social media—\(^5\) perceptions that there is a misinformation problem on social media are widespread. A recent poll illustrates the issue, showing that 82% of Americans believe they will read misleading information on social media.\(^6\) Most people (59%) think that it is hard to tell the difference between factual and misleading information, and despite years of trying to fix this problem, most people (55%) actually think this task will become more difficult leading up to the 2020 election than it was in the 2016 election.\(^7\)

Thus, blame for the misinformation problem is placed squarely on social media. Critics have therefore suggested that the solution to the problem lies with social media. If misinformation is rampant on Facebook, Twitter, Google, and other social platforms, then surely these platforms should use their technology to repair our flawed information environment. Social media platforms themselves have been eager to turn to technology—namely in the form of automated content moderation—as a means of addressing the problem.

Solutions from technology platforms generally focus on the ability of artificial intelligence and machine learning to identify and remove problematic content. This focus tends to follow the “A, B, C” model of content moderation,


\(^7\) Id.
so-called for its focus on manipulative Actors, deceptive Behaviors, and harmful Content. Different pieces of data—including information about the person sharing the post (the “actor”), the context of actions surrounding the share (the “behavior” of the poster), and the post itself (the “content”)—are input into a machine learning algorithm which is taught to classify content as either true or false. Content that is flagged as false—or likely false—is either removed from the platform entirely or demoted in such a way that fewer people actually have the opportunity to view it. The technology companies in question are, indeed, pursuing these types of solutions. Most platforms have at least some public policy about how they deal with misinformation, and most of them use some form of automation to flag, identify, and remove or de-amplify false or misleading content and actors. But here lies the problem. Quite simply, it is one of scale. Given the sheer volume of information transmitted through social media, technology for content moderation cannot possibly solve the problem alone. Nor, I will argue, might we want it to do so.

The remainder of this Paper highlights some problems with technology-driven content moderation and proposes an alternative approach to dealing with misinformation at scale on social media.

---

9 Id.; see also Limeng Cui et al., dEFEND: A System for Explainable Fake News Detection, 28 PROC. ACM INT’L CONF. ON INFO. AND KNOWLEDGE MGMT. 2961 (2019).
11 See, e.g., id.; Colin Crowell, Our Approach to Bots and Misinformation, TWITTER (June 14, 2017), https://blog.twitter.com/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html [https://perma.cc/6LPE-7HT8]. An important caveat to this has emerged in the last year, wherein at least one major player in this space—Facebook—has said it will not monitor the content of political ads by candidates running for office, on the premise that such an approach is “grounded in Facebook's fundamental belief in free expression, respect for the democratic process, and the belief that, especially in mature democracies with a free press, political speech is the most scrutinized speech there is. Just as critically, by limiting political speech we would leave people less informed about what their elected officials are saying and leave politicians less accountable for their words.” See Fact-Checking on Facebook: What Publishers Should Know, FACEBOOK (2020), https://www.facebook.com/help/publisher/182222309230722 [https://perma.cc/2WWV-FLZX].
II. PROBLEMS WITH MACHINE LEARNING-BASED CONTENT MODERATION

There are two main reasons why automated content moderation on social media is not an effective solution for dealing with misinformation in that space. The first is an empirical problem. Technology cannot effectively deal with the scale of information that exists on social media. The second is an ethical problem. Social media companies are ill-equipped, from an ethical perspective, to solve the misinformation problem with technology alone.

A. Empirical Problems with Automated Content Moderation on Social Media

Classification models are very effective for certain tasks, such as identifying easy-to-label characteristics. In certain situations they can even perform better than humans. For a complicated task like identifying misinformation, though, it is unlikely that we can expect a high rate of successful classification. After all, even humans that study misinformation for a living disagree quite a bit about what misinformation is, what its characteristic attributes are, and how to identify it. Feeding information like the source of a post, the words it includes, and responses to the post into a machine learning algorithm is likely insufficient to effectively identify whether the content is true or false, even if we could agree on what truth—or the absence of truth in the form of misinformation—is in the first place.

In general, such algorithms are considered effective if they successfully classify at a level substantially greater than chance. For example, asking an algorithm to tag chat conversations as either emotional or non-emotional gives it two possible options. Flipping a coin or choosing randomly between the two (or simply always choosing one option) would be expected to produce 50% accuracy. A study that used machine learning to perform this task resulted in 90% accuracy. The key here is not simply that 90% is relatively high, but that it is high in relation to 50%, which is the baseline expectation.

Machine learning classifiers specifically trained to identify misinformation on social media vary in their success rates. Two recent

12 See, e.g., Fariba Karimi et al, Inferring Gender From Names on the Web: A Comparative Evaluation of Gender Detection Methods, 25 PROC. INT’L CONF. COMPA NION ON WORLD WIDE WEB 53–54 (Apr. 2016) (discussing the applicability of machine learning algorithms to infer an individual’s gender based on their name).
13 See Emily K. Vraga & Leticia Bode, supra note 1 at 136.
examples achieved relatively high levels of success at this task—90.1% on a medical discussion board and a mean average precision rate of .95 on Twitter. But, these numbers also mean that at least 5% and possibly closer to 10% of problematic content on social media will fail to be identified using automated content moderation.

This is not to discount the value of automated identification of problematic content—certainly that has to be part of the solution. But when dealing with the scale of content we’re talking about on social media, even a very effective identification algorithm will fail with enormous frequency.

Social media is big. Facebook, the largest social media platform in the world, currently has 2.4 billion users—roughly one out of three people in the world. YouTube and WhatsApp also have more than one billion users each. Massive user bases create an enormous amount of content. To provide a few illustrative examples, 300 hours of video are uploaded every minute on YouTube. Every day, 60 billion texts are sent on WhatsApp, 95 million pictures are shared on Instagram, 140 million tweets are posted, and 300 million photos are uploaded to Facebook.

As an example, consider this back of the envelope calculation. There are 60 billion WhatsApp texts each day. If 8.5% of those texts are misinformation, this would result in 5.1 billion pieces of false content shared through WhatsApp each day. Given this scope, even an algorithm that was able to accurately identify misinformation at an astounding rate of 99.9% (which is quite unlikely, given the difficulty of doing so, and in comparison to the rates cited above), would leave up something in the neighborhood of 510 million pieces of false content on WhatsApp every day.

To the extent that we think the problem of misinformation is inherent in the nature of misinformation itself—that is, the problem is not just that there

17 Id.
19 Id.
20 Id.
21 This hypothetical assumes that the 8.5% rate of false misinformation sharing among people from the Guess et al. study is a constant rate that applies across platforms and to message traffic rather than on Facebook and to people. See Andrew Guess et al., supra note 5. This assumption is likely untrue, but it is the best data-based assumption we have available; the actual percentage of texts that are misinformation could be lower or higher.
is lots of inaccurate information, but if any false content is being disseminated, that necessarily undermines the information environment— the inability to successfully identify all such content is particularly problematic.

And this is to say nothing of the false positive rate—the extent to which content that is not in fact problematic is identified as such. That would result in true content getting falsely identified as misinformation, and interfere with user experience and any commitment to free speech that a platform supports.

B. Ethical Problems with Automated Content Moderation on Social Media

Technical limitations aside, there remains the question as to whether we are comfortable with technology platforms deciding what should be classified as misinformation in the first place. Platforms take a variety of policy approaches to this question, and sometimes even for different specific issues.

For example, Pinterest has decided there is sufficient evidence about the safety of vaccinations to make this a clear-cut issue when it comes to misinformation. As a result, they still allow people to post about vaccinations but do not return any user-generated information when someone searches for information about vaccinations; rather they direct users to “reliable results about immunizations from leading public health organizations, including the World Health Organization (WHO), the Centers for Disease Control (CDC), the American Academy of Pediatrics (AAP) and the WHO-established Vaccine Safety Net (VSN)”. Twitter has pursued a similar approach on the question of vaccines—search results returned for a vaccine-related query also return reliable public health information (e.g., from the United States Department of Health and Human Services)—but has a much different policy when it comes to political posts. For instance, when someone on Twitter (1) is a candidate for government office or current government official, (2) has more than 100,000 followers, and (3) has a verified account, they are automatically subject to a different set of rules, as compared with normal

22 Indeed, the fact that people perceive misinformation on social media to be a massive problem, despite the fact that data shows only 8.5% of people share misinformation on social media, suggests that this is the case—people are concerned about misinformation even when it is not widespread. See Andrew Guess et al., supra note 5; Brett Neely, supra note 6.


24 Id.

Twitter users. If such a user posts content that “may be in the public’s interest” but would otherwise violate Twitter’s rules (including those related to misinformation), the content is allowed to stay up, although with a notice about how it violates Twitter policy placed over the content.

Things get even more complicated when considering different platforms’ approaches to political advertisements, which range from banning them entirely (including Pinterest, Twitter, TikTok, LinkedIn) to limiting targeting abilities (Google/YouTube) to limiting ads to those from national candidates (Reddit).

This set of examples indicates just how challenging it can be to decide what counts as misinformation and when free speech or public speech or political speech should be prioritized over truthful speech. People often think of the vaccine example, but most misinformation is much more complicated than that. When it comes to vaccines, there is strong scientific consensus about safety and efficacy. There are clear sources of expert information to share, and these sources even tend to be non-political and have bipartisan approval (80% of both Republicans and Democrats approve of the Centers for Disease Control, for instance).

---


27 Id.


30 Public Expresses Favorable Views of a Number of Federal Agencies, Pew Research Ctr.
this clear are unfortunately rare. Indeed, for many issues, we cannot even agree who the experts are in the first place.

Because most issues lack clear guidance from experts and evidence, deciding truth is not so straightforward. And the less straightforward the issue, the less likely automated content moderation is to succeed.

III. TOWARDS A MISINFORMATION SOLUTION

Both empirical challenges and ethical issues suggest that a technology-based solution to misinformation on social media is incomplete at best. The question remains as to what other potential solutions might help solve this problem.

One option that holds a great deal of promise is person-to-person correction on social media. When someone shares misinformation, it is either seen by some portion of their social network on that platform or, depending on the platform and the individual user’s setting, seen publicly. The people viewing the misinformation therefore have an opportunity to address it. Specifically, they can reply, identifying it as misinformation and correcting the original poster accordingly. This correction can originate with ordinary users or with experts or expert organizations in the area from which the misinformation comes.

The main benefit of this correction is not actually to the original poster of the misinformation (although this type of correction also tends to reduce misperceptions). Indeed, that person may well feel threatened by being corrected in a semi-public space. Rather, the greatest benefit is what my research refers to as observational correction. All of the people on the social media platform who witness both the misinformation and the correction are likely to be affected.

(Oct. 1, 2019),
Increasingly, research bears out this promise. People who observe someone else being corrected reduce their own misperceptions on the issue. This is true across a variety of social media platforms—including Facebook, Twitter, WhatsApp, and video platforms.

Person-to-person correction also works whether the correction comes via a fact checker via the platform itself, via an expert in the field, or via another social media user. This gives a great deal of flexibility in terms of who can have impact. Given that the current usership of Facebook is 2.4 billion people—nearly a third of the world’s population—the potential exists to mobilize a veritable army of correctors.

And again, an individual correction is observed by many other people, amplifying its effect. The nature of social media networks—for instance, the average Facebook user has 338 friends, and the average Twitter user has 707 followers—means a single correction of misinformation may be seen by hundreds of other users.

---

38 See generally Emily K. Vraga et al., The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Correction on Health Misperceptions Related to Sunscreen and Skin Cancer (2020) (unpublished manuscript) (on file with author).
39 Vraga & Bode, Using Expert Sources, supra note 33.
40 Vraga & Bode, See Something, Say Something, supra note 35; Margolin, Hannack & Weber, supra 36.
41 Bode & Vraga, See Something, Say Something, supra note 35; Margolin, Hannack & Weber, supra 36.
42 Ortiz-Ospina, supra note 16.
Evidence shows and numbers suggest great promise in mobilizing experts and users to correct one another. But the problem remains as to how to convince them to actually do so. This convincing will require a shift in public opinion.

Right now, people are much more likely to blame institutions like media and technology companies than they are to hold the public responsible for solving the misinformation problem. Two recent polls\(^{45}\) found people are most likely to task the media or journalists with solving this problem.\(^{46}\) Others sharing this blame include the government\(^{47}\) and technology companies.\(^{48}\) Yet only 20% of those polled by Pew and 12% of those polled by Marist thought the public should shoulder the blame for solving the problem of misinformation on social media.\(^{49}\) The public will need to take greater ownership of the problem in order to normalize peer-to-peer correction efforts, and encourage more people to engage in them.

In addition, experts and users alike seem hesitant to engage in these person-to-person correction efforts. Experts may use social media to disseminate true information or even to debunk well-known misinformation. But they generally avoid engaging one-on-one with individuals in the way that research suggests might be most effective for other social media users to witness.\(^{50}\)

Similarly, social media users report being somewhat hesitant to engage with other users, attempting to avoid the infamous “Facebook fight.”\(^{51}\) Despite this reticence, correction does occur.\(^{52}\) A recent report from the United Kingdom shows that nearly three quarters of those who admit to sharing


\(^{46}\) 53% of people say journalists according to the Pew study; 39% say media according to the Marist study.

\(^{47}\) 12% of people according to the Pew study; 15% according to the Marist study.

\(^{48}\) 9% of people according to the Pew study; 18% according to the Marist study.

\(^{49}\) Mitchell et al, supra note 46; Neely, supra note 46.

\(^{50}\) Vraga & Bode, Using Expert Sources, supra note 33.

\(^{51}\) Emily K. Vraga et al., How Individual Sensitivities to Disagreement Shape Youth Political Expression on Facebook, 45 COMPUTERS IN HUM. BEHAV. 281, 281 (2005).

\(^{52}\) Ahmer Arif et al., A Closer Look at the Self-Correcting Crowd: Examining Corrections in Online Rumors, 2017 PROC. CONF. ON COMPUTER-SUPPORTED COOPERATIVE WORK & SOC. COMPUTING 155, 155 (2017).
“exaggerated or made up” news also report that someone corrected them for doing so, and 21% of users report engaging in such correction efforts themselves.\(^{53}\) Notably, users seem to be relatively good at recognizing misinformation: only 5% of users reported that they did not share misinformation, but were nonetheless “corrected” by someone else for sharing accurate content.\(^{54}\)

Several elements have been identified that seem to motivate people to act when they witness misinformation being shared.\(^{55}\) First, there is the perceived locus of responsibility—that is, who the user thinks is responsible for the misinformation and how they view themselves as part of the information environment. This ties back to the point about public opinion—who you view as responsible for both creating and solving the problem of misinformation affects whether you are willing to act when confronted with known misinformation. Second is the corrective objective, which asks, roughly, who or what do I hope to correct? If the user thinks about the broader information space, rather than the individual they are correcting, they may be more likely to act. Related to this is the imagined audience.\(^{56}\) When considering whether to correct, does a user think about the effect it will have on the person they are correcting, or on the broader audience that might view and benefit from that correction?

This model of decision-making when it comes to correction\(^{57}\) offers guidance for how we might motivate people to be more willing to act when they encounter misinformation on social media. Specifically, an intervention reminding people of the broader audience for the misinformation post, and for any correction of it, might encourage more users to correct one another when appropriate.

Other interventions might be necessary for public health organizations to get more involved in this space. Research must consider the barriers to such actions, and incentives for engaging in them, when it comes to public health organizations.

---

\(^{53}\) Andrew Chadwick & Cristian Vaccari, Online Civic Culture Centre, News Sharing on UK Social Media: Misinformation, Disinformation, and Correction 5, 24 (2019).

\(^{54}\) Id.

\(^{55}\) Arif et al., supra note 52.


\(^{57}\) Arif et al., supra note 52.
IV. CONCLUSION

Although technological solutions alone cannot and should not solve the problem of misinformation on social media, user- and expert-driven correction offers a data-supported means of addressing the issue that relies on individual understandings of the truth. Journalists, fact-checkers, and public health organizations might think about how they can share content that makes it easy to make use of such corrections, and technology companies and media literacy interventions should consider how to motivate people to engage with one another on the veracity of shared content on social media.
LIGHT DISINFECTS

Whitney Phillips*

CITE AS: 4 GEO. L. TECH. REV. 379 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ............................................................................................................. 379
II. IN THE BEGINNING THERE WAS LIGHT ................................................................. 381
III. THE FREEDOM OF FREEDOM .................................................................................. 383
IV. BEARING WITNESS ................................................................................................... 388
V. DIGITALLY-MEDIATED LIGHT ..................................................................................... 394
VI. ECOLOGICAL ILLUMINATION .................................................................................... 397

I. INTRODUCTION

The assertion that “sunlight is the best disinfectant,” sometimes stated as “light disinfects,” is so common on the contemporary Internet that it often goes without saying. It’s just true, no argument is needed. When someone does situate the term, the attribution invariably goes to Supreme Court Justice Louis Brandeis, who in 1913 wrote that “sunlight is said to be the best of disinfectants; electric light the most efficient policeman.”1 Although Justice Brandeis was describing how transparency and regulation can minimize financial crimes, the phrase “sunlight disinfects” resonated far outside the specific context of finance. It has since become a widespread aphorism, one that, according to The American Prospect’s Mark Schmitt, very quickly transcended insight “to cliché and beyond.”2

* Assistant Professor of Communication and Rhetorical Studies at Syracuse University with a focus on digital ethics, journalism ethics, and disinformation; Ph.D., University of Oregon, English with a folklore and digital culture focus. This article dovetails with and builds upon research conducted for my forthcoming book You Are Here: A Field Guide for Navigating Polluted Information, co-authored with Ryan M. Milner; I thank him for the feedback and insight. Thanks as well to Ezra Klein, on whose podcast I first ran up against the bifurcated nature of light. I failed to make my point clearly then. This piece represents a months-long effort to better explain what I meant.

1 LOUIS BRANDEIS, OTHER PEOPLE’S MONEY 92 (Frederick A. Stokes Co., 1914).
Reflecting on the ubiquity of the idea within the news media, *Vox* Editor-at-Large Ezra Klein links the sunlight model to journalists’ fundamental duty to inform. According to this model, we have to call attention to harmful things like mass shootings, white supremacy, and everyday presidential racism, if we hope to do anything about them.

Journalists are not the only proponents of light as a disinfectant. For example, in 2013 the Anti-Defamation League included the phrase “Sunlight is the best disinfectant” as a subject header in a press release about the growth of European anti-Semitism. Oren Segal, the director of the Anti-Defamation League’s Center on Extremism, repeated the claim in 2018 when responding to white supremacists’ efforts to recruit on college campuses. “It’s a cliché,” Segal explained in an interview with *The Washington Post*, “but we still believe that sunlight is the best disinfectant.” The assumption that light disinfects is also a common belief among everyday citizens; in my own work exploring far-right media manipulators, polluted information, and bigoted attacks online, I frequently encounter variations on the phrase as well as strong resistance to anyone who suggests otherwise.

Very often, I am a person who suggests otherwise. I whole-heartedly share the concern about white supremacy and a catastrophically polluted information ecosystem. My objection is that both things can be worsened, or at least profoundly complicated, by the potent collision of sunlight and digital tools. Algorithms that docent users to ideologically-siloed and other trending content, the post-now-ask-questions-later incentives of the attention economy, and the ease of searching for, storing, remixing, and redeploying digital content all send information zooming unchecked across audiences. Light refracts and shifts to warp speed, growing increasingly unpredictable with each network turn. While that light might disinfect for some, for others it can promote dangerous ideologies and cultivate ever-worsening threats to public health.

Put simply, the sunlight model doesn’t always do what people think it does. In fact, it can cause an even bigger mess downstream. The cliché and beyond must therefore be dethroned as a natural, universal solution to harmful speech and behavior. Instead it must be regarded—and respected—as

---


something fundamentally ambivalent, with as many potential harms as benefits. Exploring the embodied challenges of the sunlight model is the first step to this end. Doing so establishes a “digital control” that isolates which of these challenges are inherent to the light itself and which emerge from—or at least are exacerbated by—digital tools. The second step to achieving a more nuanced understanding of light is to cultivate ecological thinking about online environments: specifically, how the light we shine is reciprocally interconnected with our networks, our tools, and ourselves. Ecologically-sensitive approaches to light help minimize the harms and maximize the benefits of the deceptively simple declaration: look at this.

II. IN THE BEGINNING THERE WAS LIGHT

For many, “light disinfects” is gospel. Within the Judeo-Christian tradition, light is genesis, God’s creation of the world. Light is such a pervasive motif in Judaism, Rabbi Adin Steinsaltz explains, that “redemption, truth, justice, peace and even life itself ‘shine,’ and their revelation is expressed in terms of the revelation of light.” Christianity places an equivalent emphasis on light with an additional theological counterweight: darkness. Darkness isn’t merely an absence of light; it’s the state of being spiritually lost, of being blinded to the truth, of traveling down the wrong path without God’s guidance or grace. Darkness is, as a result, often apocalyptic. Christian eschatology hinges on light defeating dark, a feature especially characteristic of the extreme dualism of Catholicism that crystalized during the Middle Ages. This extremity is most clearly evidenced by the European witch craze from the 14th to 17th centuries, which overlapped with the Catholic Inquisition. Inquisitors and their supporters asserted that the Devil was on the march, and that the only thing that could stop him was God’s divine, righteous light.

In an ironic twist, historians Rolf Reichardt and Deborah Louise Cohen argue, the battle between good and evil, light and darkness, was so central to Christian theology that it provided the aesthetic and rhetorical framework for the Enlightenment. These lumières emerged from a Christian—and particularly French Catholic—tradition, even as they railed against that

---

8 Nachman Ben-Yehuda, The European Witch Craze of the 14th to 17th Centuries: A Sociologist’s Perspective. 86 AM. J. OF SOC. 1, 5 (1980).
tradition. The continuity between the visual symbolism of the Enlightenment and the visual symbolism of Christianity was so pronounced that Reichardt and Cohen described it as a “secular, historically updated reinterpretation of the Old Testament creation myth.” The primary difference was that the divine light of God was no longer what restored direction to the lost and sight to the blind. That’s what scientific truth was for, as darkness was reframed as a counterpoint to progress and empiricism.

Visual, philosophical, and literary representations of truth-as-light locked in battle with ignorance-as-darkness persisted throughout the centuries. The concept permeates Victor Hugo’s proclamation in Les Misérables that “[t]he true division of humanity is this: between the luminous and the dark. To diminish the number of the dark, to increase the number of the luminous, behold the aim. This is why we cry, education, knowledge!” Truth-as-light is also implicit in the claim made by Justice Louis Brandeis—who wasn’t just steeped in Western legal traditions, but also in Jewish ontology—that “sunlight is the best of disinfectants.” What he said resonated because it had the ring of hundreds of years of truth; light as a kind of magic.

That’s not the only reason “light disinfects” has persisted as a self-evident truth. While the assertion might seem straightforward, its use is bifurcated, reflecting two different sets of assumptions, stakes, and ethical paradigms. People think they’re talking about the same thing because they’re using the same words. But that same thing is actually two arguments. The resulting equivocation complicates clear-eyed analysis; it’s difficult to hold either argument up to scrutiny when each is shrouded by the other.

One of those arguments emerges from liberalism. Reflecting its deep philosophical roots, “light disinfects” in the liberal sense replicates the visual motifs of the Enlightenment, and the Enlightenment’s replication of the visual motifs of Christianity. These include images of mirrors showing things “as they are,” blazing suns, and bright horizons, all symbolizing truth. Liberal light tends to be aimed at the bad action—the hate, the abuse, the ugliness—itself. It disinfects by filtering those harms through the marketplace of ideas. This process, it’s assumed, exposes hate and falsehood for what it is: a dark cloud of ignorance. In so doing, liberal light strips hate and falsehood of its power. The second meaning and implicit argument for “light disinfects” aligns with social justice activism. It tends to focus on those affected by harmful actions. The light of social justice disinfects by inviting others to bear witness

9 See generally Reichardt & Cohen, supra note 7.
10 Id. at 95.
11 VICTOR HUGO, LES MISÉRABLES 93 (Charles Lassalle, 1863).
13 See generally Reichardt & Cohen, supra note 7.
to the affected parties’ first-person, subjective experiences of pain. The idea, and ideal, behind this light is that seeing harmful effects helps catalyze a collective process of truth and reconciliation.

Cleaving the light of liberalism from the light of social justice invites a more exacting analysis of what each light is meant to do and what it actually does, or at least, is capable of doing. Zeroing in on how each light works in a historical, offline context also helps pinpoint what elements of society are being illuminated—allowing us to ask, is there somewhere else we should be looking?

III. The Freedom of Freedom

The goals and priorities of the light of liberalism reflect the goals and priorities of liberalism more broadly, particularly its emphasis on individual autonomy and personal freedoms. As communication scholars Clifford C. Christians, John Ferré, and P. Mark Fackler explain, these freedoms are best understood as negative freedoms: freedom from external restriction. Liberalism’s emphasis on freedoms from censorship, regulation, and nanny-state handholding (or at least, perceptions of hand holding) helps to explain the overlap between liberalism as a political philosophy and conservative politics, a connection distilled as don’t tread on me.

Radical freedoms pose challenges, of course, especially when one person’s freedoms bump up against another’s. The solution within liberalism is not to restrict those freedoms, but rather to let the markets decide. Regarding speech, liberalism relies on a minimally-regulated marketplace of ideas. Informed, rational subjects exposed to a variety of speech from many quarters will weigh each position carefully and, after deliberation, embrace the best and brightest arguments—no censorship needed. The light of liberalism plays a critical role in this process. It exposes people to the full range of argumentative possibilities and, the argument goes, enriches and educates society as a whole. Despite these broad pro-social aims, the responsibility for bearing all that light, and responding logically to the lights of others, falls to individual citizens. Their individual truth-telling will out, and in outing will disinfect, and in disinfecting will preserve our personal freedoms.

---

15 Focus on markets is an especially prominent feature of neoliberalism. For more on the role markets play within the liberal tradition, see Stephen J.A. Ward, The Invention of Journalism Ethics: The Path to Objectivity and Beyond 193 (Philip J. Cercone, 1st ed. 2006).
16 See generally Christians, Ferré & Fackler, supra note 14.
American folklorist Alan Dundes provides a textbook example of the light of liberalism and its implicit reliance on the marketplace of ideas. In 1983, Dundes, writing with Thomas Hauschild, published a collection of Auschwitz jokes circulating Germany. These were not jokes told by Jews, Dundes emphasizes. These were jokes told by anti-Semites, or at least by those who thought violent anti-Semitic humor was worth telling and re-telling. The jokes proved that anti-Semitism wasn’t dead in Germany. At the same time, Dundes maintains, the jokes showed that people were talking about the Holocaust, which in Dundes’ estimation was better than not talking about it at all.

Dundes published a follow-up with Uli Linke, unsubtly titled “More Auschwitz Jokes,” in 1987. Besides presenting a fresh collection of jokes, the collection chronicles the fallout from his 1983 article. Not only did the Journal of Jewish Studies refuse to review the article, Dundes explains, but he also received a number of “hateful letters” once it was published, and many argued he never should have recirculated the jokes. Dundes defends his article by claiming that the jokes would have been circulating anyway, even suggesting that if someone had published more accounts of anti-Semitism before WWII, lives may have been saved. That was the purpose of publishing these jokes: to shine a light on Germany’s persistent anti-Semitism, not just among the extremist ranks, but as part of everyday joke-telling. Perhaps not all the joke tellers and listeners actively hated Jews, Dundes conceded, but clearly some did, and any dehumanization of Jewish people, even in the context of “just joking,” was something to address head-on. His reasoning was simple: left on its own, evil doesn’t just disappear. It needs to be aired out and revealed for what it is, and that is the job of the folklorist—to fight injustice by shining light on injustice.

Dundes then echoes the Enlightenment’s mirror motif, stating that folklore doesn’t create society, it reflects it. The ugly reality mirrored is what needs to be altered, not the existence of the mirror itself. Dundes concludes that, “[u]nless or until the causes and extent of prejudice are recognized, that prejudice will persist. To the degree that folklore is a factor in the formation and perpetuation of prejudice, it must be held up to the light of reason.” And

17 See generally Alan Dundes and Thomas Hauschild, Auschwitz Jokes, 42 Western Folklore 249 (1983).
19 Id. at 29.
20 Id.
21 Id. at 30, 31, 37.
22 Id. at 23, 38.
to what end? “Perhaps one day, Auschwitz jokes, or jokes like them, will no longer be told.”

“An Open Letter to Race Hatred,” written by William N. Robson and produced by CBS in 1943, makes a similar argument about confronting bigotry head-on and is particularly illustrative of what the light of liberalism seeks to achieve by publicizing hate.

The program opens by imploring its audience to listen carefully, because “we believe no sensible, fully informed American will allow to happen again here at home what he is fighting against all over the world.” The program’s subsequent denunciation of Hitler’s systematic attacks against Jews makes clear what sensible, fully-informed Americans should be on guard against (“the pattern is the same, the victim similar. The minority which is most easily recognized”). The stakes could not be higher, the program asserts; so long as bigotry engulfs the United States at home, the effort to fight fascism abroad will be a hypocritical fool’s errand, and a weakness the Axis powers will surely exploit.

As an example of how US racism was interfering in the war effort, the program then begins to narrate the Detroit race riots, which took place several months earlier. Dozens of Black people were killed and hundreds injured during the riots, prompting the deployment of federal troops. A fictionalized radio broadcast from Japan—which the program includes to illustrate the global impact of the riots—describes these victims as having been “sacrificed to the altar of American white superiority complex.” Flashing forward two days, the program zeros in on the graduation ceremony of a desegregated high school that had been placed under federal protection. A mob of “Kluxxers, cowards and crackpots” broke through the police line, threatening a group of white and Black students. Their bayonets raised, the troops charged; they quickly sent the rabble back “into the shadows whence they came,” reflecting the centuries-old equation of darkness and ignorance.

The announcer then pivots to the moral of these stories. Facing the riots and the self-sabotaging effects of race hatred “quietly and without passion or

---

23 Id. at 38.
25 Id.
26 Id. at 74.
27 Id. at 73–74.
28 Id. at 62–68.
29 It’s unclear if this characterization was shared by the program’s producers, or if they were speculating about how the riots would be weaponized as anti-American propaganda. Id. at 73.
30 Id.
31 Id.
prejudice”—in other words, holding a mirror up to American bigotry—lays bare a critical truth. It’s up to individual people to stop it. “It’s each one of us,” the announcer states, “each anonymous citizen keeping his head on his shoulders, his fists unclenched, and his mouth shut.” The call for individual citizens to make the reasonable, responsible choice to reject prejudice is then reiterated by a brief address from politician Wendall Wilkie, a white supporter of civil rights. White American racism, Wilkie argues, tramples the rights of Black citizens. It’s also a roadblock to maintaining good relationships with non-white war allies. It’s therefore imperative to ensure the rights of Black people at home and to guard against the “forces of fascism” that seek to deprive some Americans equal citizenship. The fascist attitude within our own borders, Wilkie argues, “is as serious a threat to freedom as is the attack without . . . It is essential that we eliminate it at home as well as abroad.”

As both examples show, the light of liberalism means well. The light of liberalism certainly makes a compelling case, particularly about fighting the forces of fascism and bigotry. The problem is that the light of liberalism easily backfires, resulting not in the clear-cut, universal disinfection of hate, but also its proliferation.

These unintended consequences stem from a series of faulty assumptions. The first is that the marketplace of ideas is a fundamentally neutral and, indeed, rational apparatus; that everyone is equally free to add their perspectives, and that individual critical thinkers weighing all the available evidence will arrive at the most rational, progressive positions, thus yielding the most rational, progressive consensus. This simply isn’t how the marketplace of ideas works. For one thing, as free speech lawyer Nabiha Syed emphasizes, it has long favored the voices and perspectives of dominant populations, namely white people and white men in particular. Non-dominant voices are either ignored or shouted down, disallowing those perspectives from ever becoming true contenders within the cultural upvoting

32 Id. at 62.
33 Id. at 75.
34 Id. at 76–77.
35 Id. at 77.
36 Id.
37 See generally CHRISTIANS, FERRÉ & FACKLER supra note 14.
process.\textsuperscript{39} The marketplace of ideas, in other words, isn’t a particularly rational, fair, or free mechanism to begin with.\textsuperscript{40}

That the marketplace is replete with short-sightedness and bias speaks to another false assumption made by the light of liberalism: that facts are enough to change people’s hearts and minds. The belief that the rational liberal subject will see a fact, apply a neutral process of critical thinking to it, then come to an objectively correct conclusion, is the underlying argument for shining as many lights as possible on as many injustices as possible. The light of liberalism thus aligns itself with the federal troops described in “An Open Letter to Race Hatred,” who, bayonets raised, send the Kluxxers, cowards, and crackpots scurrying back into the shadows whence they came.\textsuperscript{41}

This argument falls apart, of course, if the federal troops brandishing the bayonets are themselves bigots. It also falls apart when considering just how unreliable an ally the truth can be when correcting false or harmful information. Alice Marwick emphasizes this point in her critique of truth as a “magic bullet” in the fight against disinformation. I make a similar point in work exploring the limitations and unintended consequences of fact checking.\textsuperscript{42} People believe and do things not solely because of facts, but because of a host of other complex, social-psychological factors that often have very little to do with rationality.

Despite the faultiness of these assumptions, the light of liberalism remains compelling and seemingly self-evident, because the ideals of liberalism remain compelling and seemingly self-evident. The result is to provide de facto justification and even logical predicate for flooding the darkest corners of the marketplace with light—a position most simply articulated in the assertion because it works. What history shows, however, is that it does not work, at least not reliably. What the light of liberalism reliably does instead is call greater attention to bigots’ messages, making them more prominent, mainstream, and accessible—a point Dundes himself acknowledges when he admits that his piece aided in the circulation of the

\textsuperscript{39} “Upvoting” is a term popularized on the content aggregation site Reddit. Reddit allows users to click an up arrow icon beside posted content; the more users click that arrow, the higher in the feed the content travels, allowing more people to see and engage with it. Conversely, if users click the down arrow, the content falls in visibility. See generally REDDIT, http://reddit.com [https://perma.cc/TV4R-P2FF].

\textsuperscript{40} See generally CHRISTIANS, FERRÉ & FACKLER supra note 14, at 25–30, 38–44.

\textsuperscript{41} BARNOUW, supra note 24 at 73.

Holocaust jokes he published, and that before publication, most people weren’t aware they even existed.43

Early newspaper coverage of the Ku Klux Klan provides another example of how the light of liberalism can amplify and strengthen hate. As historian Elaine Frantz Parsons chronicles, Northern journalists played an enormous role in strengthening the Klan during Reconstruction.44 The Klan’s racial terrorism was all too real. But newspaper coverage lent coherence and national branding opportunities to what had been an inchoate and uncoordinated group of racist vigilantes, ensuring more opportunities for more violence.45 Historian Felix Harcourt describes a similar dynamic in the rise of the second Klan in the 1920s.46 Even the most scathing coverage, Harcourt emphasizes, helped spotlight the group for nonmembers.47 As a result, Klan membership skyrocketed. Here too, incessant publicity played an enormous role in helping the Klan secure its place as the United States’ “Invisible Empire.”48

If disinfection really were an inherent quality of light, the more coverage, the more condemnation, the more exposés of the Klan there had been, the less powerful the group would have become. The opposite proved to be true, spurring the Black press during the 20s to adopt a policy of defiant silence when responding to Klan activities—a position Jewish groups in the 1960s also urged journalists to adopt in response to the rise of the American Nazi Party.49 A minority of voices has long understood an uncomfortable truth: that not only can light not be trusted wholesale to disinfect, it risks setting in place the weapons of one’s own murder.

IV. Bearing Witness

Unlike the light of liberalism, which foregrounds the individuals composing society, the light of social justice foregrounds the society comprising individuals—a society whose unjust norms, structures, and

43 Dunb & Linke, supra note 18, at 29.
44 See generally Elaine Frantz Parsons, Ku Klux: The Birth of the Klan During Reconstruction (2015).
45 See generally id.
46 See generally Felix Harcourt, Ku Klux Kulture: America and the Klan in the 1920s (2017).
47 Id. at 19–21.
48 See generally Harcourt, supra note 46.
hierarchies must change if there is any hope for individual citizens’ actions to change. From this framework, negative freedoms are supplanted by the positive freedoms of communitarian thinking, namely freedom for the good of the collective. Ethical action, Christians, Ferré, and Fackler explain, is thus reconfigured as a “positive duty to create a social environment in which others can share the same rights equally.”

As it reflects the communitarian assertion that individuals are never alone and that their rights always already exist in relation to the broader community, the light of social justice hails the collective we, not the atomistic me. In the process, it provides an alternative to what Christians, Ferré, and Fackler describe as liberalism’s “mistaken assumption that our personal identity exists independently of socially given ends.” An ethics centered on the we first, me second—that, the authors assert, is the true disinfectant “for our moldy conventions.”

As when considering the light of liberalism, highlighting how the light of social justice unfolds in embodied contexts helps isolate the variables and challenges specific to digital environments. The death of Emmett Till is a case in point; it exemplifies the vast differences between where the light of liberalism shines and where the light of social justice shines.

In 1955, the 14-year-old Till traveled from his home in Chicago to Money, Mississippi, to visit family. While there, he visited a white-owned store, bought something, and left. The store owner’s wife Carolyn Bryant, with whom Till had briefly interacted, later accused Till of sexually threatening her—an accusation Bryant recanted in 2017. On Carolyn Bryant’s word, Till was hunted down and lynched by store owner Roy Bryant and several accomplices; his mutilated body was later discovered in the Tallahatchie river, weighed down by a cotton-gin fan. Till’s mother, Mamie Till Bradley, insisted that her son’s body be transported back to Chicago. Journalists chronicled the arrival of Emmett Till’s casket at the train station, its transfer into the hearse, and departure for the funeral home. After seeing her son on the slab, Mamie Till Bradley made a decision: the press must also chronicle her son’s broken, mangled body. “Let the people see what they did to my boy,”

---

50 See generally CHRISTIANS, FERRÉ & FACKLER, supra note 14.
51 Id. at 45–46.
52 Id. at 46–47.
55 GORN, supra note 53 at 1, 32.
56 Id at 56–63.
57 Id.
she famously said, and allowed David Jackson of *Jet* magazine to take photographs.\(^{58}\) *Jet* published the images in September 1955, which were then reprinted by a number of Black newspapers, including the *Chicago Defender*.\(^{59}\)

According to popular history, the image of Till’s mutilated body galvanized white Americans, helping to shore up national support for civil rights. This is a compelling story, historian Elliot Gorn explains, but is complicated by the fact that, until the 1980s, very few white people had seen David Jackson’s funeral-slab series.\(^{60}\) Eventually, Mamie Till’s wishes would be realized; the people, including the white majority, would see what the bigots did to her boy (many of them misremembering when they first encountered the photos). But initially, the most gruesome images of Emmett Till were limited in circulation—though they were not limited in their effects, successfully galvanizing a generation of Black Americans towards civil rights activism.\(^{61}\)

The images of Till’s funeral, however, and Mamie Till’s excruciated expression as she so publicly grieved, were widely circulated within the white press.\(^{62}\) The result, political scientist Heather Pool argues, was to center the struggles of marginalized people for sympathetic whites, unify heterogeneous groups through collective grief, and lay bare the failures of the United States’ democratic ideals.\(^{63}\) What Mamie Till was seeking, and what she successfully achieved, was for her son’s fellow citizens to ask the question, “why was this boy allowed to die?”—a question whose answer could generate precisely the collective energies needed to enact structural change.

In insisting that the country—and indeed the world—bear witness to her son’s violent murder, Mamie Till aligned herself with other powerful civil rights activists working long before the Civil Rights era. The post-Reconstruction anti-lynching campaign of Ida B. Wells, for instance, similarly implored: *you will look at this.* Wells was a Black journalist who began her career in Memphis, Tennessee.\(^{64}\) When three of her friends were lynched in 1892, she resolved to speak truth to white supremacist power.\(^{65}\) When she did so, racist whites threatened to kill her; Wells moved north and kept writing.\(^{66}\)

\(^{58}\) *Id.* at 59, 62.

\(^{59}\) *Id.* at 62.

\(^{60}\) *Id.* at 2, 62.


\(^{62}\) *Id.* at 1–2, 7.


\(^{65}\) *Id.*

\(^{66}\) *Id.*
In three pamphlets published between 1892 and 1900—*Southern Horrors: Lynch Law in All Its Phases, Red Record, and Mob Rule in New Orleans*—Wells dismantled the cover stories used to justify lynchings. These attacks were not, Wells showed, isolated events unconnected to race or racism, as many white Southerners maintained. They certainly were not honorable efforts to protect white women, a racist myth that positioned white Southern men as heroes, not murderers. Lynchings were, instead, acts of racial terrorism and fit within a broader systematic effort to control Black people and restrict power to white men. This was an affront to American values, Wells asserted; the truth had to be told. “When the Christian world knows the alarming growth and extent of outlawry in our land,” she wrote in *Red Record*, “some means will be found to stop it”.

This call, at least in verbiage, was similar to the one made in “An Open Letter to Race Hatred.” However, rather than framing the solutions in terms of the individual citizen, Wells sought to “intervene boldly in public discourse and to change public opinion so that the application of justice for all could prevail,” as biographer Jaqueline Jones Royster explains. Wells focused particularly on educating white Americans, who had the ability—and therefore, according to Wells, the obligation—to push for change. Ultimately, Congress failed to pass the federal anti-lynching laws for which Wells and others campaigned. Over a century later, they still haven’t. Still, Wells forced critical debates on the subject of lynching, and public sentiment against mob violence did indeed shift. In shining her light, Wells created the conditions for a different type of individual to emerge: one aligned with the communitarian freedoms of a more just society.

The Equal Justice Initiative’s Legacy Museum in Montgomery, Alabama continues in Wells’ footsteps. It tells the story of slavery from the perspective of those harmed, dehumanized, and murdered through

---

67 *Id.*
68 *Id.* at 27–33.
69 *Id.* at 28.
70 *Id.* at 157.
71 *Id.* at 40.
72 *Id.*
73 *Id.*
75 See generally EQUAL JUST. INITIATIVE, [http://www.eji.org](http://www.eji.org) [https://perma.cc/U3A2-NC88].
enslavement, and shows how the institution of slavery evolved from lynchings to Jim Crow to mass incarceration. The museum grounds are also home to the National Memorial for Peace and Justice, which includes an installation of hundreds of suspended columns representing the counties in the United States where lynchings occurred. Each of the columns bears the names of those who were murdered. The museum’s guiding ethos is that shining a light on ugly truths is a necessary first step towards restorative justice. The museum’s executive director Bryan Stevenson emphasizes this point, arguing that the United States never fully contended with the full ugliness of slavery or the defiant and widespread white resistance to the civil rights struggle. Because the root of that evil was never confronted, it has been allowed to grow and evolve. Appropriately, the Legacy Museum’s museum book includes a full-page quote from Ida B. Wells: “The way to right wrongs is to turn the light of truth upon them.” The Legacy Museum thus invites visitors to bear witness and confront America’s racist past with “honesty, courage, and hope for the future.” Only then can the country begin the process of collective healing. Only then can freedom really be free for all.

The light of social justice does not presume that the marketplace of ideas will filter the best ideas to the top; in fact it implicitly concedes that the most resonant and popular ideas are often the ones in most desperate need of challenging. Lynching, for example, was popular. These rituals of pornographic torture and murder were enthusiastically attended by thousands of white Southerners. White newspapers chronicled every move, often covering the white terrorists sympathetically. White children looked on, unfazed. The marketplace of ideas was, at best, useless to the cause of social justice, and at worst, an accomplice to mob violence.

The light of social justice, in short, reveals the unseen—including the myopias of liberal light. It contends with power. It shows us bodies. It is necessary to an inclusive, pluralistic democracy. But it is not a failsafe. It too can have unpredictable effects. It too can backfire. In her reflections on the pain of others, Susan Sontag forwards a simple explanation: the fracture of the

---

77 Id.
79 Id. at 5.
80 Id. at 2.
“we” who looks. “No ‘we’ should be taken for granted when the subject is looking at other people’s pain,” she writes.\textsuperscript{82}

For audiences grounded in the same ethical paradigms as those shining the light, social justice spotlights can serve as what visual culture scholar Ariella Azoulay describes as an emergency claim.\textsuperscript{83} Emergency claims highlight something so egregious and so extraordinary that it triggers a pressing sense of civic responsibility to act.\textsuperscript{84} \textit{How terrible, this is not the kind of society we want; we must do something to stop it.} When successfully made and received, emergency claims can catalyze meaningful social change.

The wild card is who encounters those claims, and what insular “we” those audiences identify with. Sometimes (certain) audiences only register the first part: that this thing is terrible, and doesn’t reflect the kind of society we want. And yet that recognition is then followed by awkward silence. Condemnation without follow-through easily slips into abstraction. Literary scholar Debra Walker King describes this risk in her analysis of \textit{blackpain}, the process by which Black bodies in distress are transformed from spiritually-whole subjects to metaphors for suffering.\textsuperscript{85} Those observing these bodies “from the trees,” that is to say, with a comfortable emotional distance, can cry about something sad that happened to (the idea of) a Black person, feel good about crying, then continue on with the day as if nothing happened.\textsuperscript{86} The essence of \textit{blackpain} is an emergency claim that serves as nothing more than a set-piece for education, or even entertainment—a particular risk once that pain is made consumable by white people.

Sometimes, both moral revulsion and a sense of social responsibility are lacking: in short, both halves of the emergency claim fail. For that segment of the “we,” an event is not regarded as all that bad, so there is no reason to do anything about it. In these cases, Azoulay explains, harms are normalized as the status quo.\textsuperscript{87} For many white people in the North and South, this was lynching in a nutshell. As political communication scholar Richard Perloff chronicles, white newspaper coverage embodied such an outlook.\textsuperscript{88} Reporting on lynchings was “akin to reporting on unpleasant acts of nature such as earthquakes or floods; the events were unfortunate but necessary aspects of the order of things...”\textsuperscript{89}

\textsuperscript{82} \textsc{Susan Sontag}, \textit{Regarding the Pain of Others} 7 (2006).
\textsuperscript{83} Ariella Azoulay, \textit{The Civil Contract of Photography} 197 (2008).
\textsuperscript{84} Id. at 197–99.
\textsuperscript{85} \textsc{Debra Walker King}, \textit{African Americans and the Culture of Pain} 16–17 (2008).
\textsuperscript{86} Id. at 20.
\textsuperscript{87} \textsc{Azoulay}, supra note 83, at 203–04.
\textsuperscript{88} Perloff, supra note 81, at 317–321.
\textsuperscript{89} Id. at 318.
Speaking to the outcry following Emmett Till’s death, Heather Pool offers up an even more distressing outcome of a failed emergency claim: that members of the dominant group double-down on the status quo. \(90\) This is particularly risky when the emergency claim is directed towards the group responsible for the harms in question. Many white Southerners, for example, reacted to Till’s death not as a catalyst to racial reconciliation, but as further reason to mistrust outsiders and activist groups like the NAACP, who these white people believed were conspiring to destroy the “traditional” (read: insular, homogenous, and structurally racist) Southern way of life. \(91\) And that’s to say nothing of the white people who were compelled to raise the Confederate flag, whether proverbial or literal, even higher. This second, more explicitly violent response is evidenced by how frequently Emmett Till memorials have been shot up over the decades. Indeed, so many people have shot at the Graball Landing memorial sign—which marks where Emmett Till’s mangled body was fished from the Tallahatchie River—since its dedication in 2008 that the Emmett Till Memorial Commission had to replace it four times. The commission finally installed a bulletproof sign in 2019. \(92\)

When considering the consequences of failed emergency claims, the takeaway is stark: Where there is apathy, the light of social justice cannot shine as intended. Where there is hate, it can transform into something else entirely. The more unwieldy the audience is, the more difficult it is to predict which response is more likely. For example, given that visitors could be traveling in from any part of the country for any range of reasons, it is not possible to anticipate whether the people driving towards the Graball Landing memorial are there to confront the United States’ violent past with honesty, courage, and hope for the future—or whether they are there to riddle the memorial with bullets.

V. Digitally-Mediated Light

Online, predicting an audience’s response—let alone identifying where one audience ends and another begins—can be even more difficult. The permeability of audiences online stems from social sharing spurred by trending topic algorithms, streamlined retweeting and reposting functions, and the various attention-economy incentives dangled by profit-driven social

---

90 Pool, supra note 63, at 418.
91 For more on how the Till case was reframed as an outside attack against white Southerners, see Gorn, supra note 53, at 64–68.
platforms. Information often moves faster than it can be tracked in real-time, especially as it weaves between public and private networks. Anticipating who will see something we post, and what might happen as a result—or, put another way, who will show up to mourn and who will show up with guns—is often impossible to predict.

The consequences for light, both of liberalism and social justice, are profound. A person trying to call attention to something might shine their light nobly, a steady beam cutting through the darkness. However pure that light might be, however focused, its reception is in fact prismatic, refracting wildly across networks. Its colors change; its wavelengths lengthen; and it can never be called back.

Because it holds a mirror up to society to reveal its ugliest contours, the light of liberalism is particularly vulnerable to out-of-control refraction online. The funhouse mirror that is the marketplace of ideas only strengthens that light, bending it towards worse and stranger outcomes. This happens because there is no singular, self-contained marketplace online. Instead, groups are sorted into highly-insular echo-systems—a term introduced by Kate Starbird and her team at the University of Washington—93—that keep people algorithmically fed by a steady diet of worldview-affirming media. This is particularly true of echo-systems on the right, which over a period of decades have become asymmetrically polarized, even radicalized.94 These echo-systems essentially function as highly-specialized marketplaces of ideas. A scathing critique spotlighted by the light of liberalism can filter into a reactionary marketplace and emerge transformed: into a joke, a justification, or an incentive to do something even worse next time.

The possibility that the light of liberalism might reconfigure into something dangerous speaks to a further vexing fact of light: that calling attention to harm, even in order to minimize that harm, still serves to publicize the harm. The ambivalent risks associated with publicity have long been an Achilles heel of the light of liberalism. Online, the unchecked spread of information, coupled with hopelessly collapsed audiences, only exacerbates the problem. I explore the digital contours of this tension in my 2018 Oxygen of Amplification report on best practices for reporting on bigots, manipulators,

---

and abusers.\textsuperscript{95} When asked about the dangers of amplifying hate by reporting on hate, one journalist encapsulated the issue. As he explained, “There’s no way around [the publicity tension]...There’s bad people in the world, and there are poisonous ideologies in the world, and at a certain point you have to realize that you’re promoting them to a…”\textsuperscript{96} The reporter paused. “Not promoting them, but you’re getting those ideas out to a wider audience.”\textsuperscript{97}

Almost by journalistic definition, getting those ideas out to wider audiences risks handing weapons to those audiences. Of course—and this is where the reporters I interviewed furrowed their brows most deeply—not taking that risk could be worse. Remaining silent could allow the problems to fester unchecked, give bad actors control of the narrative, or signal complicity. The challenges truly cut both ways.

Digital spaces pose similar complications for the light of social justice. As necessary as the light might be to communitarian freedom and justice, online its potential benefits are often matched by its potential fallout. For example, the Equal Justice Initiative—whose Legacy Museum and National Memorial for Peace and Justice provides visitors a brilliantly curated, immersive, and viscerally interactive experience—disabled the comment feature on all of its promotional YouTube videos.\textsuperscript{98} The purpose of the museum is to foster dialogue and engagement. Online, that learning environment is much more difficult to control. The unwanted actions of unintended audiences—pressingly, the violently racist YouTube comments those audiences might post—is simply not worth the risk. Such commentary could, and almost assuredly would, transform the videos into sites of white racial terror.

In short, context-collapsed audiences online pose significant challenges to social justice efforts. That’s not all: these same challenges are easily weaponized by bigots. Even when a person highlighting harms takes great care to honor the experiences of those subjected to hate, harassment, and violence, bigots can take the most thoughtful framing and flatten the person or people described into racist, easily sharable, memes. In other words, into content—an abstraction that very easily dovetails into Debra Walker King’s

\textsuperscript{97} Id.  
conception of black pain. Sharing stories about abuse and harassment—even when the goal is to elicit a collective, communitarian response—can also spur worse abuse and harassment. Once again, the tools of digital media are the culprit. Search indexing, which allows people to easily find people, places, or things by keyword, sets those things into place, transforming them into static, easily accessible targets. This danger extends well beyond the person who has been targeted; the very light that seeks to highlight injustice in order to fight injustice lights up that person’s social networks, providing bigots and abusers a whole new set of people to terrorize.

VI. ECOLOGICAL ILLUMINATION

The light of liberalism and the light of social justice are different. Each should be analyzed on its own terms. Simultaneously, digital spaces blur that dividing line. For one thing, it can be extremely difficult to tell one kind of light from the other once it’s been refracted through seamless social sharing. For another, both kinds of light can cause considerable, if wholly unintended, harms.

In these cases, the most dangerous quality of light is that it does not seem dangerous. We assume that light is inherently good; that it’s the best of disinfectants. It is true that light can be good and can disinfect in some cases, with some people. But light can also be the stuff of nightmares. Exploring the history of light—including all it has solved and all it has made worse—helps dispel the idea that we can take our light to the bank. All we can take to the bank is refraction.

This does not mean we are powerless. It certainly does not mean we should default to a defeatist binary: either we never shine our light, or we shine it regardless of consequence. Both will only make things worse. What we must do, instead, is approach light ecologically.

Ecological thinking draws lessons from the interconnection and interdependence of the natural world and applies those lessons to digital spaces. For example, in nature, the deep reciprocities between flora and fauna, predator and prey, sky and earth, ensure that no one thing can be carved out from all the rest. Everything is dependent upon everything else; what happens to one happens to all. The Internet is the same. There are deep reciprocities between our digital tools—our algorithms, our retweet buttons, our media editing software—and our everyday interactions with those tools (algorithms, for example, feed us, but we also feed them). There are deep reciprocities between the bigots, chaos agents, and manipulators (collectively,

“bad actors”) and citizens of good faith; the actions of one shape and incentivize the actions of the other. There are deep reciprocities between our sprawling institutions and the body politic, both of which circulate, respond to, and strengthen good and bad information alike.

An ecological approach to light explores these connections by asking how the things we illuminate here might impact the people, places, and things over there—directing our attention to the unexpected places and unintended consequences of everyday action. An ecological approach to light also sidesteps questions of intentionality. We might want to rid the world of the scourge of bigotry; we might want to protect and empower marginalized groups; we might want to defend democracy with all our might. And yet, however deeply we care about our communities and the environment, intentions are not the same as outcomes, especially online, when information zooms unpredictably across and between countless networks with little ability to predict what happens next.

By asking us to reflect on how, when, and where we shine our light, ecological thinking reminds us to ground ourselves in the world we’re actually in, not the one built atop existing assumptions. Living bravely in that world means applying strategic, case-by-case assessments of the costs and benefits of light and of darkness. It means acknowledging that our light can be weapons for some, and necessary beacons for others. Most importantly, it means acknowledging the impact we have on the world and making peace with the fact that, when it comes to light, there are no easy answers.
HOW THE INTERNET CREATED MULTIPLE PUBLICS

Lam Thuy Vo*

CITE AS: 4 GEO. L. TECH. REV. 399 (2020)

TABLE OF CONTENTS

I. INTRODUCTION .......................................................... 399
II. INFORMATION SEGREGATION THROUGH ALGORITHMS .......... 400
III. ALGORITHMIC SELECTION AND DISTRIBUTION OF INFORMATION ...... 402
IV. CONSTANTLY CONTEXTUALIZED INFORMATION AND POLITICALLY PERFORMATIVE CONTENT CONSUMPTION ........................................ 408
V. A SHIFT OF WHAT IS POLITICALLY ACCEPTABLE .......................... 410

I. INTRODUCTION

Political identification in the digital age has shifted online: increasingly, people define their political identity in how they come together around issues and news events on the social web.

We adopt online political identities in three major ways: through shared consumption of information on social media platforms; through participation in political movements through hashtags and around news events; or through performance of our political identity via virtue signaling on the Internet. From the alt-right to Bernie bros, online communities coalesce around news articles and other information that allows them to express their political affiliations through the content they read, react to, and share. And through this consumption of similar information, they form little political information universes often referred to as “media ecosystems.”

* Lam Thuy Vo is a senior reporter at BuzzFeed News where she digs into data to examine how systems and policies affect individuals. She has an expertise in mining and analyzing data from the social web and is currently a visiting researcher for the Technology and Social Change Project at the Harvard Kennedy School's Shorenstein Center on Media, Politics and Public Policy. In 2019, she published a book, Mining Social Media, about her empirical approach to examining data from the Internet with No Starch Press.

1 See Carolyn E. Schmitt, ‘Network Propaganda’ Explored, HARVARD GAZETTE (Oct. 25,
These universes are segregated in the kind of information they consume due to the ways in which the social web is engineered. The Internet caters information to people in highly personalized ways and often delivers more of the same through algorithms rather than serendipitously. It is optimized for the virality of one-punch headlines, not stories with nuance. And this pushes political information universes further apart than they may otherwise naturally be.

Thus, we face a fractured political landscape online—multiple publics, if you will, that are increasingly informing our real-life affiliations, too. And with that, we are also contending with multiple realities and various spectrums of what is politically acceptable to different online groups, when we look at politics as a nation.

This Paper will closely examine group formation around information online through the lens of one platform: Facebook. While this Paper cannot comprehensively study political coalescing online, it offers one example of the effect of online information consumption on political identity formation and the segregation of information universes brought about by this form of online assembling. While Facebook is just one of many social media platforms that play a part in how political identity and political groups are formed in the digital age, analysis of this one platform is enough to shine a broader light on the kinds of issues affecting the political landscape today.

II. INFORMATION SEGREGATION THROUGH ALGORITHMS

While political identity is shaped by a myriad of factors that exist in the physical world—such as the political leanings of one’s geographical milieu, familial connections, or class-related experiences—a lot of the content we consume on social media platforms has become an increasingly important basis for people’s understanding of politics and thus for their relationship to it. A study of the impact of social media on the 2016 presidential election from Ohio University, for instance, noted that more people cited Facebook as a source for political information than any other news-related site.2

The ways in which people consume information on social media platforms like Facebook, however, are very different from traditional media models. Unlike those models where a select number of editors and reporters gather, package, and publish information that consumers will read or consume on their own or in small groups, information on social media is delivered in

---

highly personalized ways that favor polarizing content. Consuming content on social media may, thus, potentially exacerbate existing political divides. And, as information universes become more segregated, the consumption of articles, posts, and visual content has become a nucleus around which people politically coalesce.

In a way, information consumption is both skewed towards consumers’ natural tendencies by virtue of whom people are friends with and what pages they “like,” and increasingly through performative acts that online consumers use to signal their political affiliations within these information universes.

Thus, social media users express their politics while consuming it. This can include actions like sharing specific articles from partisan outlets or using verbiage and humor that is oftentimes specific to a person’s political information universe. Think for instance of the word “snowflake” that is often used by conservative groups to insult people who identify as liberal for being “too sensitive.” Using these words or sharing content that may contain these words is a clear demarcation of oneself politically and serves as a way to self-identify to spectators in semi-public spaces like Facebook.

Political actions like this abound on the social web. From political protesting through hashtags, to sharing articles on Twitter around a specific subject, to the intense discussions in the comment sections of a Facebook post — the social web has turned the Internet into a metaphorical public plaza on which people gather to show their political stripes.

To better understand how information segregation and the resulting politically performative way of consuming, sharing, and opining on information affects social structures and political factioning, it is first important to understand how social media platforms have changed the context in which we consume information:

1. The information people consume is selected algorithmically and based on human, often emotional, reactions to information, rather than editorial selection, resulting in highly personalized information environments, often also referred to as “filter bubbles.”

2. Information on the social web is always surrounded by emotional responses from others and designed in a way that encourages participation through emotional reactions, commenting, or sharing. Content is, therefore, automatically contextualized both emotionally and within the social framework of “other people,” with the public space turning content consumption into a performative act.

---

III. ALGORITHMIC SELECTION AND DISTRIBUTION OF INFORMATION

What surfaces on the timelines of our social media feeds is largely determined by the data that social media companies collect from consumers’ behavior on their platforms.

First, there is the element of the self-selection of information that determines what data is even available in the personalized realm of consumers’ information universe. Social media feeds can only contain content produced by the media outlets, publications, and human beings who were selected into the personal information universe of each social media user. On Facebook, self-selection takes two forms: the people one chooses to be “friends” with, as well as the groups and pages that one chooses to include in their interests. The company itself has referred to the content available to surface on one’s timeline as “inventory.”

Then, the social media companies may use different data points and algorithms to evaluate the content from this inventory. Data points that may be included could be the point in time when a piece of content was published (and in particular its recency to the point in time when a timeline is accessed); how fast one’s Internet is; and also, and perhaps more importantly, how people who exist in the information universe of each individual have reacted to or interacted with a piece. Facebook refers to these data points as “signals.”

Then, social media platforms use algorithmic decision-making to surface different kinds of content on one’s timeline. For Facebook, the process of “ordering” content onto any individual’s newsfeed employs algorithmic predictions based on previous data, which can include algorithmically deduced scores on how likely a user is to comment or to hide a story.

While Facebook does not publish its algorithmic ranking mechanism, the company explicitly stated in a press release from January 11, 2018 that their algorithmic ranking of information prioritizes posts “that spark conversations and meaningful interactions between people.”

As the company states:

To do this, we will predict which posts you might want to interact with your friends about, and show these posts higher in

---

6 Id.
feed. These are posts that inspire back-and-forth discussion in the comments and posts that you might want to share and react to—whether that’s a post from a friend seeking advice, a friend asking for recommendations for a trip, or a news article or video prompting lots of discussion.

We will also prioritize posts from friends and family over public content, consistent with our News Feed values.\(^8\)

In a later update, the company confirmed that this policy is still intact:

“\textquote{The News Feed algorithms prioritize posts that are predicted to spark conversations among people, whether because of format — for example, live videos tend to lead to more discussions than regular videos — or because the posts were shared by people, groups or Pages you interact with frequently.}”\(^9\)

As Meredith Broussard explained in her book, *Artificial Unintelligence: How Computers Misunderstand the World*, a basic understanding of an algorithm is “a computational procedure for deriving a result.”\(^10\) Algorithms often rely on data to evaluate, predict, or make decisions, according to Broussard.\(^11\)

Given Facebook’s prioritizing of posts that “spark conversations,” a lot of the political information that people encounter is informed by decision-making algorithms that likely rely at least in part on the ways in which people emotionally react to posts or how much they partake in the online discussions within the comments of a post.

What often seems to remain unexamined when discussing these decision-making algorithms on Facebook are the *kinds* of data sets that are being used to measure what kind of information sparks conversations. As Rashida Richardson, Jason Schultz, and Kate Crawford argue in their paper, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, biased data may skew algorithmic


\(^11\) Id. at 94.
decision-making in ways that replicate biases in policing.\textsuperscript{12} They find that policing algorithms that rely “on data produced during documented periods of flawed, racially biased, and sometimes unlawful practices and policies . . . cannot escape the legacies of the unlawful or biased policing practices that they are built on.”\textsuperscript{13}

Similarly, it is important to question the kind of data that is used to define what content “sparks conversations,” according to Facebook. While the company has not made public how exactly their algorithms work and what data is used to determine their rankings, there are clues in what kind of longitudinal user data is collected. There are the kinds of data points collected through people’s interactions with posts on their news feeds (reaction buttons such as “haha,” “wow,” “love,” “like,” “angry,” and “sad;” the “hide post” button; the comments; etc.).

While the data collected from these kinds of actions—from emotional reactions to comments—does not seem biased at first, it may be important to consider what is not captured in the data collection of social media behavior to truly understand the limitations of the data that powers the algorithms of the timelines of these platforms.

On Facebook, for instance, classification of emotional reactions is limited to the aforementioned six categories, which arguably capture emotions only in their extremes and may omit other, less tangible signals. These less tangible signals may prioritize sparking thoughts and nuance, rather than “conversation” (which often becomes argumentation) and stark emotional reaction. As artist Mimi Onuoha pointed out through her 2016 art installation, \textit{The Library of Missing Data Sets}, we need to consider missing data points as a counterpoint to what is measured:

“Missing data sets” are the blank spots that exist in spaces that are otherwise data-saturated. Wherever large amounts of data are collected, there are often empty spaces where no data live. The word “missing” is inherently normative. It implies both a lack and an ought: something does not exist, but it should. That which should be somewhere is not in its expected place; an established system is disrupted by distinct absence. That which we ignore reveals more than what we give our attention to. It’s in these things that we find cultural and colloquial hints of what


\textsuperscript{13} Id. at 204.
is deemed important. Spots that we've left blank reveal our hidden social biases and indifferences.\textsuperscript{14}

Within Onuoha’s framework, we should understand the data that feeds algorithms in their decision-making within the context of not just of what \textit{is} measured, but more importantly also of what may \textit{not} be measured. For example, comments and reactions may only capture reactions of some of the most vocal consumers of information or may only capture reactions that respond to some of the most provoking and less nuanced content. They may also omit responses to content that either requires more processing time from a person or is not represented in the options that a social media platform’s interface offers.\textsuperscript{15}

There is also good reason to believe that these “extreme” reactions come from a small but vocal fraction of people who actually consume content on social media. This gives disproportionate influence on the content of a news feed to active and vocal consumers, something that I have called ‘the tyranny of the loudest.’ For instance, in an experiment published on BuzzFeed News, I measured the number of reactions a video live stream on two partisan Facebook pages, Fusion and Fox News, received against the number of views the platform displayed for the same video.\textsuperscript{16} Only two to three percent of the number of people who viewed the videos actually decided to react or comment on either of the two streams.


Therefore, algorithmic selection of the information that populates the feeds of individuals is heavily skewed towards posts that elicit strong emotional reactions or a large number of comments. This potentially exacerbates existing differences and further segregates information universes instead of giving equal balance to nuanced and moderate voices and those on the outer ends of emotional spectrums.

It is difficult to consider this argument in a holistic, empirical way due to the lack of transparency provided by social media companies. However, one experiment conducted for BuzzFeed News may illustrate how algorithmically powered “filter bubbles” can exacerbate existing political differences. The story captured the divergent experiences of information online of a conservative mother and her liberal daughter. These two people care about each other deeply but believe in different political values. They said that their consumption of political information online led to conflicts that they could have otherwise worked through in person.17

An analysis of 2,367 posts on their newsfeeds showed how their information universes were deeply shaped by the politics of both the outlets they followed, but more importantly, by the people who made up their social circles.

Both subjects told BuzzFeed News that their experiences of political content online was deeply divisive to them and that, when discussing the same issues in person, they were able to speak to one another in a more nuanced way and resolve their issues. This sentiment is further amplified by reports that, in light of a divisive election, people have begun to unfriend or block various people who may oppose their political views, further limiting the inventory of content that may surface on one’s timeline.\textsuperscript{18}

Thus, it is very plausible that algorithmically-curated timelines are causing people to consume increasingly polarized content that segregates them from their political counterparts.

IV. CONSTANTLY CONTEXTUALIZED INFORMATION AND POLITICALLY PERFORMATIVE CONTENT CONSUMPTION

While algorithmic decision-making skews the inventory of content that people encounter online towards partisan (if not hyperpartisan) content, it is people’s conduct online that truly turns the consumption of this information into a performance of political identity.

The social web is designed to encourage both participation around information and, to a lesser degree, the actual consumption of and deep engagement with that information. Seldom are passive metrics, like content views or time spent with content, displayed.\(^{19}\)

It is a somewhat commonly held belief that a large proportion of people share articles and information without having thoroughly engaged with them (often failing to read beyond a headline). There are several product designers who have pointed out that the way that a lot of the social web is designed does not allow for people to thoroughly engage with the content they encounter on social media.

In a Netflix episode of Abstract, for instance, an Instagram product designer expressed remorse about bringing the “endless scroll” to viewers, saying that this format does not leave users with time to digest the content they consume.

Similarly, the inventor of the Retweet button on Twitter told BuzzFeed News writer Alex Kantrowitz that he regrets inventing the tool because it made it much easier for users to share content. Before the button’s invention, people had to copy and paste messages into their own status bars before ‘retweeting’ it:

“Copying and pasting made people look at what they shared, and think about it, at least for a moment. When the retweet button debuted, that friction diminished. Impulse superseded the at-least-minimal degree of thoughtfulness once baked into sharing.”\(^{20}\)

---

\(^{19}\) On Facebook and other platforms, views are often only displayed for live or one-off videos.

Some design features of online platforms hence encourage quick participation on online platforms, rather than more thoughtful critical engagement.

To some degree, then, the social web—as it is engineered and as has been proclaimed online many times—is where nuance goes to die.

Users’ actions are also seen by a semi-public audience, making these actions always somewhat performative. These performative actions then become a way in which the content that surfaces online is framed. Information on the social web is always surrounded by emotional responses from others and hence will, to some degree, require users not just to take in and process the content at hand, but also to take sides. The question faced by users ultimately becomes: Do I agree with the predominant emotion that has been generated by this piece of content or not? The other option is to opt out of participating and to consume silently, but even without political participation, opting out while remaining a consumer of content may exacerbate the effect that the loudest voices have on skewing what kind of content appears on people’s social media timelines.

This performativity is particularly pronounced both around polarizing news events and when people coalesce around prominent political personalities, demonstrating just how personal politics has become for online users.

For example, in a previous article about Alexandria Ocasio-Cortez, I looked at the fandom and anti-fandom practices around the highly visible congresswoman, and showed how the use and remixing of memes about her can be seen as political acts demarcating the contours of political factions:

. . . Virality, meme culture, and fandom are interwoven with how we view and understand politicians. As politics has crept more and more into our timelines and Instagram feeds, it’s also arguably become more personal than ever. Not everyone is obsessed with #AOC, but media cycles and viral social posts sure make it seem as if we should be.

The internet has turned politics into cliques that mimic a lot of high school dynamics: Each clique comes with its own lingo, its own way of celebrating its heroes or putting down its opponents.

Do we side with the popular girl? Do we demonize her and show the world just how “stupid” she is by bullying her? Or do
we sit in the corner of the cafeteria and watch it all silently, unclear as to where we belong?21

V. A Shift of What is Politically Acceptable

Given these two conditions—the skewing of political content users see on social media and the increasing performativity of consumption online—and given the clustering of Internet communities around these polarized political information universes, it is perhaps useful to take a closer look at what ideas are acceptable within these universes.

The concept of the Overton Window has recently received a lot of media attention in light of U.S. President Donald Trump’s rhetoric and has been a helpful framework for people to understand how the U.S.’s “political imagination” has shifted under Trump’s presidency.22 The concept may also be helpful in examining the political filter bubbles that exist online.

The conservative think tank that originated the term defines it as a:

model for understanding how ideas in society change over time and influence politics. The core concept is that politicians are limited in what policy ideas they can support— they generally only pursue policies that are widely accepted throughout society as legitimate policy options. These policies lie inside the Overton Window. Other policy ideas exist, but politicians risk losing popular support if they champion these ideas. These policies lie outside the Overton Window.23

While this concept is often applied to society as a whole, it is important to note that information segregation and performative information consumption have potentially also created political landscapes with groups that each contain their own Overton Windows. If political groups are increasingly consuming different sets of facts and are inhabiting these universes in ever more finely cut factions, it may be logical to assume that those consumers also have varied understandings of what is a politically defensible policy or behavior.

Recent reports around the impeachment hearings, for instance, have pointed out that Trump supporters may be inhabiting worlds of information that are almost entirely divorced from information presented by the witnesses.\textsuperscript{24} As Washington Post reporter Isaac Stanley-Becker, who examined the comments of Trump supporters from a private Facebook group, phrases it:

“Trump’s most ardent supporters have fashioned alternative realities for themselves—as well as for Republican lawmakers aiming to turn the charge of corruption back on those investigating the president.”\textsuperscript{25}

In this instance, online political coalescing around information has created a different understanding of what is and is not constitutional behavior by a president.

To better understand the political edges of the Overton Window in any given political online group, it may be helpful to observe the infighting among group members. By snarkily or humorously expressing dislike of the political leanings of a piece of content or of a political or public figure’s actions (a process sometimes described as “dunking”\textsuperscript{26}), or by celebrating and amplifying other political messages, online users within each political online faction are helping consumers understand the edges of their Overton Windows.

Given that most online users have curated their online social milieus to consist of people and institutions who are more likely to echo their own points of view, understanding these edges becomes even more important. The edges of each information universe are thus defined by people who are already in similar political camps.

An article about how supporters of presidential candidate Bernie Sanders attacked candidate Elizabeth Warren online by posting emojis in her Twitter replies en masse and running campaigns to undermine her may help illuminate just how narrow the Overton Window is for what is politically


\textsuperscript{25} Id.

acceptable for this group of online users.\textsuperscript{27} Both Sanders and Warren were arguably considered the most progressive among the Democratic presidential candidates at the time and, due to their proximity in political affiliation, may have been more prone to attack one another online, where information universes are deeply connected to existing political alikeness. It is, for example, less likely for Sanders supporters to encounter deeply conservative information online and to join an online group “dunking” of Mitt Romney supporters than it is for those Sanders supporters to do so to Warren supporters, due to the afore explained information segregation.

Similarly, there is evidence of infighting within far-right movements on college campuses\textsuperscript{28} that speaks of a fairly narrow Overton Window in the far-right information universe.

Multiple value systems have previously co-existed in society, but people’s increasing reliance on social media for political information has exacerbated previous trends:

- **Information segregation**: With much of our information distribution systems relying on automated mechanisms, many online consumers have lost exposure to the information that others consume and thus, of one another’s realities.
- **Consuming information is a political performance**: The way we consume information on social media always encourages us to react to it emotionally in semi-public settings, rather than to internalize it on our own. Thus, the way we interact with content signals our political affiliations to others.
- **Parallel realities**: Many online users are now building values based on a common set of facts, a trend that has clearly also started affecting the kinds of Overton Windows that politicians can move within as well.


VOTER SUPPRESSION & VOTER TURNOUT

Our second panel explored new strategies for technologically enabled election interference, including automated strategies for depressing turnout from targeted voter groups and attempts to redesign the 2020 census and change how potential voters are counted. This panel also considered new uses of networked information technologies to increase voter turnout along with the pitfalls and unintended adverse effects of such techniques.

Professor Julie Cohen of the Georgetown University Law Center lead this panel. She was joined by Cara Brumfield of the Georgetown Center on Poverty and Inequality, Jessica Huseman of ProPublica, Kathryn Peters co-founder of Democracy Works, and Nick Monaco of the Institute for the Future.

In the following pages, we are pleased to present articles by Ms. Brumfield and her co-author, Jae June, Ms. Peters, and Mr. Monaco and his co-author, Sam Woolley. We have also included Chris Conrad’s Student Note, which deals directly with voter suppression, in this section.
THE RISKS AND REWARDS OF CONDUCTING A CENSUS IN THE DIGITAL AGE

Cara Brumfield* & Jae June Lee**

CITE AS: 4 GEO. L. TECH. REV. 415 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 415
II. THE ONLINE RESPONSE OPTION ........................................... 419
III. MODERN ANALYTICAL METHODS AND BIG DATA .................. 422
IV. MISINFORMATION AND DISINFORMATION IN THE NETWORKED ERA .. 424
V. GETTING OUT THE COUNT ................................................... 426
VI. CONCLUSION ....................................................................... 427

I. INTRODUCTION

Every decade, the U.S. Census Bureau undertakes its constitutional mandate to count every person once, only once, and in the right place.¹ The 2020 Census is the nation’s first “digital” decennial census; for the first time in its history, the Bureau has invited the majority of the public to participate using an online self-response portal.² Along with the new Internet self-response option, the Bureau has pioneered and adopted cutting-edge

** Jae June Lee, Policy Analyst, GCPI ESOI, B.A. Kenyon College. The authors would like to acknowledge Indivar Dutta-Gupta and Maria Filippelli who provided insightful feedback on the article; Sheila Naughton who provided excellent research assistance; and the Georgetown Law Technology Review for their helpful edits. Any errors of fact or interpretation remain the authors’.
² In addition to the online self-response option, the public has opportunities to respond by phone or mail.
methodologies and technologies that may help contain overall costs while providing significant gains in efficiency of the 2020 Census and the quality of the published statistics.³

While every decennial census is a vast and staggeringly complex undertaking, census taking in the digital age presents particular challenges and risks to achieving a fair and accurate count. These include the “digital divide” (i.e. the demographic gaps in access to and use of the Internet),⁴ cybersecurity risks,⁵ the potential for the use of commercial and third-party data sets to “de-anonymize” (or reconstruct and reidentify) census respondents,⁶ and the threat of disinformation campaigns reminiscent of those during the 2016 U.S. presidential elections.⁷ Some experts fear the Bureau did not do enough to mitigate and manage those risks in the lead up to the 2020 Census.⁸

Amidst these tech-related challenges and risks, the COVID-19 pandemic has also created obstacles that are unprecedented both in nature and magnitude, causing the Bureau to adjust census operations and request


revisions to statutory reporting deadlines. Notably, the Bureau delayed many of the critical in-person field operations designed to reach households with unreliable Internet access and households that have not self-responded to the 2020 Census. These new circumstances have highlighted the benefits of the Bureau’s new design for the 2020 Census—such as the online self-response option and the use of big data—while also increasing the Bureau’s reliance on them, placing additional pressure on the 2020 Census’s innovations. Yet, the Bureau remains committed to its goal to “count everyone once, only once, and in the right place”.

The stakes for getting the 2020 Census right could not be higher. The decennial census is the statistical backbone of our nation and its fairness and accuracy is a civil rights issue at the heart of our democracy. Census data is used for the apportionment of seats in the House of Representatives and redistricting at all levels of government; the census is essential to determining the appropriate allocation of power within our political system. Census data is also used for the enforcement of provisions in the Voting Rights Act and other laws that help protect against discrimination. When groups are not fully counted in the census, not only is their political power diluted, but their communities may not get their fair share of funding for important programs that support economic security, health, education, and more. In fact, the allocation of over $1.5 trillion dollars in federal funds is guided by census

---


data. Put simply, when the count is not fair and accurate, the groups who are undercounted lose out the most.

That is why outreach to “hard to count” groups is so important. Hard to count groups are those who have been historically undercounted in the decennial census, such as people of color, people with disabilities, people experiencing homelessness, renters, recent and/or undocumented immigrants, and people displaced by disasters. The Census Bureau finds certain groups harder to count for a number of reasons, including language barriers, distrust of the federal government, and housing units which are hard to find.

Census experts have raised both questions about the effectiveness of the Bureau’s outreach campaigns to encourage participation among people of color, and doubts about the Bureau’s ability to hire enough culturally and linguistically competent enumerators to follow up with households who do not self-respond. The NAACP filed a lawsuit claiming that census underfunding would “result in a massive and differential undercount of communities of color.” and that “[s]uch a dramatic undercount will especially dilute the votes

---


17 Hard-to-Count Communities in the 2020 Census, supra note 15.


of racial and ethnic minorities, deprive their communities of critical federal funds, and undervalue their voices and interests in the political arena.”

Fortunately, there is also a groundswell of creative, research-based, and effective campaigning to Get Out The Count (GOTC). National and local organizations are promoting census participation by raising awareness and helping the public understand its importance—and benefiting from social media and other tech-based approaches in order to do so. Trusted messengers who know their communities have an important role to play in encouraging census participation.

Will the efforts of the Census Bureau and the vast network of GOTC advocates be enough to ensure a fair and accurate census count this year? Only time will tell.

II. THE ONLINE RESPONSE OPTION

The Bureau has offered—for the first time—an option for households to self-respond to the census using an online self-response portal. As a result of declining overall response rates for surveys and censuses and the complexity of counting an increasingly large and diverse population, the costs of undertaking the decennial census have grown significantly over the past decades. In an effort to control these escalating operational costs and make participation in the census more convenient, the Bureau introduced the new Internet self-response option—alongside other innovations—as a key feature of the 2020 Census architecture that is designed to cost less per household than the 2010 Census. Starting mid-March 2020, most households received an invitation in the mail with instructions for how to respond to the

---

25 U.S. CENSUS BUREAU, supra note 3.
2020 Census. Households may also choose to respond by phone or by mail. Households that have not self-responded will receive several reminders from the Census Bureau.

Although the online response option will make responding to the census easier and more convenient for many people, hard to count communities are among those likely to have limited Internet access and those more vulnerable to hacks and malware. For example, individuals who lack access to a reliable Internet connection at home may also have limited digital literacy and rely on older or public devices with less robust protections against online threats.

Census Bureau staff and census stakeholders have been wary of a major website failure similar to the flawed launch of the healthcare.gov website—which would dramatically interrupt operations and reduce self-response. Australia’s first census to be conducted primarily online experienced a major failure when the website used to collect responses went down after a Distributed Denial of Service (DDoS) attack. To minimize the risks of such a failure, the Bureau consulted and worked with government advisors from the US Digital Services and 18F, an office within the General Services Administration. The Bureau has taken additional steps to address concerns, when possible, or otherwise manage risks identified by the Government Accountability Office and others.

Due to “social distancing” measures adopted by state and local authorities to slow the spread of the COVID-19 pandemic, the Census

---


28 Id.

29 Id.


32 Brown, supra note 8.

33 U.S. GOV’T ACCOUNTABILITY OFFICE, supra note 19.

34 CTRS. FOR DISEASE CONTROL & PREVENTION, INTERIM US GUIDANCE FOR RISK
Bureau delayed and revised census operations that require in-person visits to households—and placed greater emphasis on the online response option. Some of the affected census operations are designed to reach communities with limited access to a reliable broadband connection. For example, the Non-Response Follow Up (NRFU) operation, which involves sending census workers door to door to collect responses from households that did not self-respond, has been delayed. The Update Enumerate operation involves sending census workers to remote areas of the country to enumerate people in person, and that operation has also been delayed. These operations are essential to the success of the 2020 Census, and these delays mean that the Bureau must rely even more heavily on self-response.

As of May 15, 2020, 59.3% of all U.S. households have responded to the census, according to the Bureau’s numbers that are reported live to the public. Most households—about 80.9% of reported responses—have responded online. These figures do not include individuals and families who live in certain group living arrangements (referred to as group quarters) and are counted through a separate operation. In the 2010 Census, a final total of

---


39 Id. (indicating the national Internet self-response was 48%; this is 80.9% of the national self-response rate of the 59.3% national response rate).

66.5% of households self-responded to the census—the Bureau expects around 60.5% of households to self-respond to the 2020 Census.

III. MODERN ANALYTICAL METHODS AND BIG DATA

Over the past decade, the barriers to using big data have significantly decreased with the advent of cloud computing and modern analytical methods. These advances have allowed the Bureau to explore new opportunities for improving the coverage, quality, and cost-effectiveness of the decennial census. Over the past two decades, the Census Bureau has employed a particular source of big data—federal and state administrative records.

Administrative records refer to information collected by government agencies and commercial entities for the purposes of administering programs and providing services, such as tax and medical records. Administrative records are distinct from data collected for statistical purposes through traditional surveys and programs such as the American Community Survey or the decennial census. For the 2020 Census, the Bureau will use administrative records to help determine the best times of the day to contact households, identify vacant housing units, and substitute missing data. The Bureau will also employ federal and state administrative records to model and produce statistics on citizenship.

While the COVID-19 pandemic has disrupted census operations, administrative records, as with the online-response option, may be even more important. However, key sources of administrative data have themselves

44 Id.
45 U.S. CENSUS BUREAU, supra note 3.
47 Informational Release, U.S. Census Bureau, supra note 9.
48 Kenneth Prewitt, supra note 11.
been impacted by the pandemic. For example, the Bureau relies on data from the Individual Tax Returns and Informational Returns provided by the Internal Revenue Service (IRS) to improve operations and to statistically estimate an individual’s missing characteristics. However, the tax filing deadline has been delayed due to the pandemic.\textsuperscript{49} While information from past filings are available to the Bureau, the Bureau may not receive complete data for the most recent fiscal year when it had planned to.

There are important limitations to the utility of administrative records, even in the best of times. For some variables, such as taxable income (which the decennial form does not ask about) or age, administrative records will likely be higher quality than self-reported data. However, for characteristics such as race (which the decennial does ask about), using administrative records may reflect the less reliable judgment of an observer rather than an individual’s identity.\textsuperscript{50}

Administrative records provide better information coverage for certain social and demographic groups over others. For example, it is likely that those who are disconnected from the labor force, people with minimal touch points with public and private record-keeping institutions, and noncitizens are among those who are less likely to be captured accurately, or at all, in these data sources.\textsuperscript{51}

The growing abundance of personal information available online and from commercial providers, along with advances in computer science and statistical techniques, present an evolving threat to confidentiality: reconstruction and reidentification attacks.\textsuperscript{52} As a steward of the public’s information, the Bureau faces a fundamental challenge in making statistics available; the more statistics that are published, and the closer those statistics match the underlying, confidential data, the greater the probability that an


\textsuperscript{51} See id.

actor can reconstruct the underlying data and identify individual respondents.\textsuperscript{53}

Reconstruction and reidentification attacks present real threats to the Bureau’s legal obligation to keep respondents’ information confidential.\textsuperscript{54} In past decades, the Bureau has employed a collection of methods, including techniques such as swapping and cell suppression, to prevent the disclosure of confidential information.\textsuperscript{55} However, Census Bureau researchers have shown that these approaches are no longer sufficient. Using publicly available statistics from the 2010 Census, Census Bureau researchers exactly reconstructed the individual records—i.e. the age, race, sex, and ethnicity of individual census respondents and the census block in which they lived—of 46\% of the U.S. population (or about 142 million individuals).\textsuperscript{56} In matching this reconstructed 2010 Census data with commercially available data, the Bureau accurately reidentified 17\% of the public, or some 52 million people, by name.\textsuperscript{57} Based on these findings, the Bureau concluded that disclosure avoidance methods used in past censuses will not be sufficient for the 2020 Census.\textsuperscript{58} This is why, for the 2020 Census, the Bureau has turned to a new mathematical standard, called differential privacy, to provide robust and measurable guarantees of confidentiality in this age of big data.\textsuperscript{59}

IV. MISINFORMATION AND DISINFORMATION IN THE NETWORKED ERA

To count everyone, every census relies on participation from the public as an important exercise of civic duty. High levels of public participation in the decennial census help the Bureau contain operational costs and improve

\textsuperscript{53} Id.
\textsuperscript{55} See JAE JUNE LEE & CARA BRUMFIELD, supra note 52.
\textsuperscript{57} Id.
\textsuperscript{59} See JAE JUNE LEE & CARA BRUMFIELD, supra note 52.
the quality of the statistics it publishes.\textsuperscript{60} By seeking to dissuade communities from participating in the census, disinformation presents a significant threat. The proliferation and ubiquity of social media provides fertile ground for the spread of mis- and disinformation about the census.\textsuperscript{61} Given the greater reliance on self-response due to the COVID-19 pandemic, the ability of the Census Bureau, community advocates, and tech and social media platforms to combat misinformation and disinformation will be even more important.

Community advocates, community-based organizations, and other GOTC organizations are working together to identify and respond to disinformation on social media.\textsuperscript{62} As trusted messengers within the community, GOTC organizations are crucial to inoculating communities against mis- and disinformation by circulating accurate content about the census.\textsuperscript{63}

GOTC organizations are working closely with the Census Bureau and information and communications technology companies. The Census Bureau has a “Trust & Safety” team to manage mis- and disinformation and a dedicated webpage and email address where the public can report any suspected misinformation and disinformation.\textsuperscript{64} Social media and technology platforms did not have Terms of Services related to the census in the past, so GOTC organizations and the Census Bureau have been working with tech companies to get census-specific policies in place.\textsuperscript{65} This includes companies such as Facebook, Google, YouTube, Nextdoor, and Pinterest. Some tech companies have since updated their policies regarding mis- and disinformation in advance of the census.\textsuperscript{66} For example, Facebook has agreed to take down misleading and inaccurate advertisements, and the Census Bureau became the first government agency to access YouTube’s trusted flagger program.

\textsuperscript{60} U.S. CENSUS BUREAU, Executive Summary, 2020 CENSUS LIFE-CYCLE COST ESTIMATE VERSION 2.0 (June 10, 2019), https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/life-cycle-cost-estimate_v2.pdf [https://perma.cc/X7J6-BMK7].


\textsuperscript{63} See id.

\textsuperscript{64} Jarmin, supra, note 7.


\textsuperscript{66} Id.
allowing them to flag videos with malicious intent (and then YouTube decides whether it violates their policies). The Bureau also has access to Twitter’s “partner portal.”

V. GETTING OUT THE COUNT

Amidst the challenges of counting in the digital age and falling trust in the federal government, community-based and national civil society organizations have an increasingly important role to play. These organizations have invested in outreach to groups the Bureau considers hard to count—and are helping fill gaps resulting from insufficient federal investment.

Groups have launched social media campaigns to target low income families and communities of color. In particular, Asian Americans Advancing Justice, NALEO, and Color of Change are building positive awareness about the census and its importance, conducting public education and GOTC campaigns such as “Yalla Count Me In,” “Queer the Census,” “Indian Country Counts,” and “Hágase Contar.” Groups are providing toolkits and workshops online for digital organizing that cover best practices and share resources among GOTC groups. Some stakeholders, like State Voices and MoveOn, have been exploring better ways to use text messages to promote census awareness and participation.

Stakeholder groups are also working to expand access across the digital divide and to leverage data and analytics to get out the count. Public libraries and other community-based organizations were planning to set up kiosks and dedicated spaces for Internet access for the purpose of filling out the census form which will both help with awareness and improve access for people who do want to respond online but might not have the Internet at home—though these have likely been impacted by the COVID-19 crisis. To

67 Id.
68 Id.
69 FULLENWIDER ET AL., supra note 62.
73 Digital Organizing Trends, Tips and Tools for GOTC Outreach, supra note 22.
74 FULLENWIDER ET AL., supra note 62.
help community advocates and community-based organizations target their outreach efforts, the Center for Urban Research at CUNY Graduate Center, has developed a map of hard-to-count census tracts and will be monitoring and reporting on real-time response rates during the enumeration.77

The COVID-19 pandemic has interrupted outreach efforts by nonprofits, businesses, local governments and civic groups that the Census Bureau had planned to rely on to help boost census participation.78 Campaigns like the Leadership Conference Education Fund’s Census Counts campaign, for example, have needed to re-strategize, prioritizing paid media and potential partnerships with essential services.79

VI. CONCLUSION

It is critical that we get the census right and ensure that all our communities are fairly and accurately counted. The rapidly growing availability of new data sources and digital technologies have presented new opportunities and challenges. This has forced the Bureau to continually adapt and change. In the strides the Bureau has taken in designing and implementing a census for the 21st century, the agency continues a long legacy of innovation. It is clear that technology and cutting-edge methodologies have an important role to play in the census—as they always have.

The same will be true moving into the 2030 census. It is impossible to fully anticipate what challenges and risks we will face, but the goal will remain the same—to ensure a fair and accurate count. As the 2020 Census continues to unfold, we will see how well we’ve managed to step up to the current challenges, and there will be lessons to carry forward. This exercise has already begun, as discussions about how to transform the Census Bureau into a true 21st century data agency—which will mean increasing reliance on new technologies and data sources—are already underway.

79 Id.
I. INTRODUCTION

Picture Election Day. It is probably a Tuesday (though not necessarily). Check-in tables in the high-school gymnasium sort voters by precinct, while plastic privacy screens fill the engine house of a nearby fire station. Some voters return ballots to drop boxes, while others mailed theirs back days ago.

Voting remains a manual process in an increasingly automated world. It is a tactile connection to democracy that goes beyond the enduringly popular “I voted” stickers. With a little squinting, one can imagine that the earliest voters would still recognize the process, in spite of the secret ballots and orderly quiet.
Look longer, and you may begin to see the increasing role of technology in elections. Many of the voters in line registered online. Others saw a reminder on Facebook or searched for where to vote using Google. And the check-in process relies on tablet-based electronic pollbooks, which connect to the local voter registration system in real time. It may be many years before we vote from our phones, but we already live in a digital democracy.

Today, technology plays three critical roles in supporting voter participation. First, it can be the medium for inviting new voters into the process. From there, it can provide consistent, accessible election information to voters. Finally, it can modernize the voting process, providing a better user experience and creating positive feedback loops for participation. However, the same tools that provide these benefits can also serve to misinform, disinform, and suppress votes. By highlighting examples of positive voter engagement and taking lessons from what works, we can define how and where technology is used in elections to support voters and build a more robust democracy.

Lessons from these examples suggest three consistent requirements for any technology used in an electoral context. As a baseline, any technology serving voters must ensure accuracy, consistency, and accessibility. Where any of these principles are not taken into consideration, new tech runs risks of irrelevance, undermining civic trust, or even disenfranchising those it seeks to help. Where they inform technology and its usage, these requirements support a more modern, inclusive democracy.

II. INVITING

“Candidates, public officials, and journalists operate in a narrow professional world that is largely of their own making and that is remote from the world of the public they serve.”

- Thomas Patterson, The Vanishing Voter

If you want someone to do something, it helps to ask them. Inviting potential voters to register is especially important, and where technology scales personal outreach, it can bring new voters into the democratic process.
Whether in comparison to other established democracies or to our own historical elections, modern American voter turnout levels are low. One group of non-voters, who Thomas Patterson labels “the disenchanted” and Kate Krontiris dubbed “interested bystanders,” follow current events and take some interest in politics, but rarely take direct part. These individuals largely associate politics with conflict and negativity, and feel unwelcome in policy conversations. Left alone, they are unlikely to take part in elections—these bystanders need positive motivation to turn out.

In 2015, I conducted a series of research interviews into the voter registration experience. My questions were all about that process: were you online? On paper? At the Department of Motor Vehicles? And yet my notes surprised me: the answers were all about people. I heard answers that began with “My dad sat me down on my 18th birthday,” or “my high school guidance counselor passed out forms.” One interviewee told me that “I saw a canvasser on campus, and he looked sad, so I went over to register and cheer him up.”

Through those interviews, I came to appreciate how much the act of invitation mattered to these voters. But for many potential voters, that invitation never comes: a 2016 Pew Charitable Trusts survey found that 62% of respondents were never asked to register to vote.

Elections can feel exclusive and unfamiliar, and being invited in (especially by a trusted entity), is key. Technology can support these invitations at scale, helping both individuals and institutions champion voter participation within their communities.

One avenue where technology has supported inviting new voters to take part is social media. As early as 2012, Facebook offered its users the opportunity to register to vote. As early as 2012, Facebook offered its users the ability to share their voter registration status as an event in their profile.

---

4 Patterson, supra note 1, at 84 (“They are disenchanted rather than alienated. They are not fuming mad at government, and unlike the alienated, they tend to believe that government has an interest in their opinions and their welfare. But they are disenchanted with how politics are conducted”).
5 Kate Krontiris et al., Understanding America’s Interested Bystander: A Complicated Relationship with Civic Duty, GOOGLE RESEARCH (June 2015), https://research.google/pubs/pub44180/ [https://perma.cc/9GXH-KLFJ].
In 2018, however, Facebook rolled out a feature that “lets people ask their friends to join them in registering to vote.”

Whereas in previous years, the reminder to register came from Facebook itself (with a social component displaying which of a user’s friends had shared their own registration status), now individual users could create voter registration posts and add their own message to the request, then track how many friends they’d reached.

No breakdowns distinguish how many voters registered specifically via a friend’s post instead of the general Facebook promotion, but the company celebrated registering at least two million voters in 2018.

On most social media platforms, users aren’t “friends” with the tool itself—with the notable exception of Snapchat. Due in part to the visual-messaging nature of the platform, Snapchat regularly distributes mass video messages to users to introduce new features and commemorate holidays. “Team Snapchat” is the first contact in any new user’s Snapchat account, and the team members quickly become familiar faces.

For National Voter Registration Day (NVRD) 2018, users received a Snap from Team Snapchat. The brief video highlighted the speed and convenience of voter registration, featuring the team’s familiar faces.

Snapchat reached more than 418,000 voters through its NVRD campaign. By comparison, a Taylor Swift PSA that ran at approximately the same time led to roughly 166,000 registrations. In both cases, an online invitation was the nudge that many people needed to get registered or update their voter registration record.

Another space where technology has expanded voter engagement is higher education. The Higher Education Act of 1998 requires that colleges and

---

10 Sandberg, supra note 8.
universities offer voter registration opportunities to students.14 Many institutions go far above and beyond this requirement, embedding civic engagement and citizenship into their educational mission. On many campuses, this outreach remains manual work, whether through staffing registration tables in high-traffic areas, planning events, or coordinating student-led programs. Where institutions have incorporated technology into their engagement plans, these tactics have helped extend their efforts and reach even more of their student population.

For example, the Harvard university registrar’s office integrated a voter registration step into their class registration process.15 Each semester, every student planning their course schedule also sees a reminder to update their voter registration or subscribe to election reminders.16 At many other institutions, Campus Labs, a vendor of student engagement software, offers a voter registration widget as an optional component in its platform.17

Initiatives like these make it possible for higher education administrators to invite every eligible student to register and vote—even across multiple campuses, commuter populations, or where other logistical factors make students difficult-to-reach with in-person events. Civic features don’t replace traditional voter engagement programs, but do help ensure that administrators can reach every student and offer them the same information and support.

Low voter participation is a persistent and challenging problem in American democracy. Non-voters are disproportionately younger, lower-income, Hispanic and Asian-American, and their omission leaves democracy less representative than it could be.18 Where technology can help to reach non-voters and invite them to take part, that simple request makes a big difference.

---

III. INFORMING

Registering is only one step in the voting process. Once invited in, voters (especially new voters) often need help navigating what’s on the ballot, when and where to vote, and what to bring. Unsurprisingly, most voters look for these answers online, and rely on technology platforms as publishers and distributors of basic election information. Initiatives that bring useful election information to voters online help to meet those needs.

These information needs are well-documented, thanks to research by the Center for Civic Design (CCD). CCD has published two voter journey maps illustrating some of the common paths that voters take from initial awareness to casting a ballot: a “happy path” and a “burdened path.” Voters on both paths seek out information about the candidates and issues, their voting options, and voting locations. And in their Field Guide to Ensuring Voter Intent, Vol. 7: Designing election department websites, CCD notes that the five most frequently-asked questions by voters are:

- What is on the ballot?
- How do I get an absentee ballot and when is it due?
- Where do I vote?
- Who is in office now?
- How do I register to vote?

Official election websites play an important role in answering these questions, but cannot do so alone. In 2008, the Pew Center on the States published “Being Online is Not Enough,” a report assessing the functionality and shortcomings of state election websites. Researchers considered how effectively election websites answered the following questions:

- Am I registered; or, how do I register?
- Where do I vote?
- What candidates and issues are on the ballot?

The researchers found that half of states offered an online voter registration status lookup tool. Two-thirds offered a polling-place lookup tool. The average usability rating scored a 58 out of 100, with a top score of only

---

In 2016, MIT’s Election Data + Science Lab’s Elections Performance Index (EPI) revisited these questions and found significant improvement, but there is still incomplete availability of voting information on official state sites across the United States.22

Alongside its findings on the availability of voting information on state election websites, “Being Online is Not Enough” announced the creation of the Voting Information Project (VIP), a collaboration between the Pew Charitable Trusts, Google, and state election officials to collect and standardize polling place data for broader online distribution.23

Since 2008, the Voting Information Project (VIP) has collected data from 44 states and the District of Columbia, and served hundreds of millions of API queries for polling place and ballot information. States provide data linking addresses to polling locations, early voting sites, and ballot information in a standard format.24 The VIP team (since relocated to Democracy Works) provides quality assurance review, then publishes the data through Google’s Civic Information API.25

By making official state data broadly available, VIP ensures that no matter where a would-be voter goes to seek out election information, they can find accurate, reliable data. The project team defines a standard data specification, provides ongoing technical support for election offices, and manages a community of developers who integrate the data into a wide variety of civic and political tools. Voters then encounter VIP data through candidate websites and outreach, on social media, or even through an SMS lookup via the GO-VOTE shortcode (46-8683).26

In addition to making official voting information more accessible, VIP has come to serve an important security role. As election officials grapple with increasingly complex threats, VIP’s data is recognized as a valuable failsafe for state election websites, even where states publish this same information directly. During Wisconsin’s 2020 spring primary, the MyVote Wisconsin tool encountered memory issues and was intermittently down for part of

23 Being Online is Not Enough, supra note 21, at 20.
26 VIP Projects, VOTING INFORMATION PROJECT, https://www.votinginfoproject.org/projects [https://perma.cc/JD6N-KGM7].
Election Day. The state promoted VIP’s Get to the Polls tool as a backup option for voters to minimize the disruption.  

Beyond VIP, Google has made additional investments in informing voters through its search products. Before an election, searches for information about candidates or voting mechanics return highlighted results that use the company’s Knowledge Panel and OneBox tools. These results address frequently-searched questions like “what to bring to vote?”  

As with VIP data, these highlighted results offer structured information, with sections like “who can vote” and “online voter registration.” Unlike VIP, this data is standardized by a national non-profit partner, Democracy Works, rather than sourced directly from election officials. The data is not published via the Civic Information API for use in third-party tools, but instead appears only in Google’s own search results.  

This standardized formatting allows Google to ensure that every voter can find useful responses in “easily digestible formats… at the top of the Search results page.”  

For many voters, casting a ballot is a complicated process. Reliable information about both the choices at stake and the steps involved helps would-be voters follow through on their intention. Election officials’ own digital presence provides a foundation that technology providers can amplify and supplement. Search platforms can prioritize clear, actionable information in results, while outreach technologies can bring information to voters proactively. These resources help move voters from a “burdened path” to a “happy path” as they navigate the many choices involved.

IV. MODERNIZING

Voter-facing campaigns to invite and inform are the most visible uses of technology, but the most important gains are found in modernizing election systems themselves. Improving technology for election administration can help register voters, increase participation, and offer voters more options for where and how they cast their ballots (e.g., early voting and voting by mail). These improvements to the voter experience help increase voter participation.

Basic measures such as database improvements can greatly improve voter registration rates. In 2002, the Help America Vote Act (HAVA) required that states establish centralized voter registration databases.  


implementation of HAVA, voter registration records were maintained by a variety of state and local officials. Though a centralized database lacks the appeal of more cutting-edge technology, this mandate established a common technological baseline that supports more visible changes to the voting process. From this deceptively simple foundation, states have built better systems for maintaining voter registration lists, automated the voter registration process, simplified support for early voting and vote-center options, and improved accountability for voting by mail.

Despite their significant impact, building databases may escape attention as a major development in modernizing elections. The implementation and design of these tools is typically an administrative activity with few publicly-visible bells or whistles. But these updates to streamline voter registration processing have registered more new voters than any public outreach campaign, and the additional security they provide reinforces public trust at a time when faith in elections and their outcomes is critical to maintain.

In January 2016, Oregon rolled out a change to its voter registration process. Where Oregonians were previously asked if they wanted to register to vote at the Department of Motor Vehicles (DMV), now they were automatically registered to vote. These new voters receive a postcard in the mail that they can fill out and return if they would like to opt out (or declare a party affiliation). Since then, a total of nineteen states and D.C. have implemented similar systems to automate voter registration, whether at the DMV or other state agencies (in Alaska, the integration is with the state’s Permanent Fund Dividend).

This shift from an “opt-in” request to an “opt-out” relies on the electronic transfer of voter registration records from state agencies to election officials. While these integrations have not always launched smoothly,

---


automatic voter registration (AVR) represents an important advance in supporting higher turnout and broader voter participation.

The federal Election Assistance Commission’s 2018 Election Administration and Voting Survey (EAVS) found “an increase of almost 10 million in the number of registration applications received by states via the DMV for the 2018 general elections compared with 2016, and the percentage of the total registrations received by the DMV (which usually processes automatic registrations) has increased by 11 percentage points compared to 2016.”

Oregon alone reported a 20-point increase in the share of voter registrations processed through the DMV between 2016 and 2018, while in Alaska, just under 60 percent of all voter registrations processed in the 2018 election cycle came through the integration with the Permanent Fund Dividend. A report by the Brennan Center estimated that AVR increased the number of registrants in implementing by jurisdictions by rates of 9 percent - 94 percent.

Given the recent implementation of many state AVR processes, their impact on turnout is not yet fully understood. An early study based on Oregon’s 2016 turnout found that 43.6 percent of automatically-registered voters voted, as compared to 84.1 percent of traditionally-registered voters. Despite this gap, automatically-registered voters made up 4.7 percent of the state’s voters in the 2016 election, marking a significant increase in participation.

Other technologies improve the accuracy and quality of voter registration records. The Electronic Registration Information Center (ERIC) is a nonprofit organization that improves the accuracy of state voter files and increases voter registration rates. By comparing voter registration records from its thirty member states and the District of Columbia and the Social Security Death Master list, ERIC provides reports on voters who have likely moved, died, or changed in eligibility status, and identifies unregistered eligible citizens within the population. Member states commit to using these reports to conduct proactive outreach based on these reports, and give those

---

34 U.S. ELECTION ASSISTANCE COMM’N, supra note 30, at 47.
35 Id.
38 Electronic Registration Information Center, ERIC Introduction, YOUTUBE (Mar. 11, 2015), https://www.youtube.com/watch?v=O8ISoeO1hjw [https://perma.cc/A4UC-D9SP].
residents additional opportunities to register to vote or update their voter registration.  

Since its inception in 2012, ERIC identified more than 34 million Americans who were potentially eligible to vote but unregistered. Each of those people received a state contact (typically a postcard) offering them an opportunity to register. A 2013 study in Washington State found that a message of ease and convenience led to a 5 percent registration rate among recipients, at a cost (in 2015) of 23 cents per postcard sent. Other states have seen response rates as high as 20 percent.

In that same period, ERIC identified millions of inaccurate individual voter registrations to member states. As with eligible-but-unregistered records, these voters receive outreach from state election officials with the opportunity to update their records or confirm that they were no longer eligible to vote in that state.

In 2016, Colorado—a founding member of ERIC—had the highest percentage of eligible, registered voters in the country. In speaking with the *New York Times*, Judd Choate, Colorado’s Director of Elections, credited ERIC as a big part of the state’s high registration rate. “ERIC has been a game changer in elections for those of us in it.”

New election technologies also support a wider range of voting methods, including early and mail voting. These offerings give voters greater flexibility in how to cast their ballots. In the 2018 midterm election, 25.8% of voters cast their votes by mail, a fraction that has trended consistently upwards over the past 30 years. Many states have implemented election laws that allow any registered voter to cast their ballot by mail.

---

43 Id.
44 U.S. ELECTION ASSISTANCE COMM’N, supra note 30, at 11.
Mail voting options have been demonstrated to increase turnout. It also offers voters additional choice and convenience and reduces the potential for long lines and other Election Day problems at polling places. New technology tools improve the security and trustworthiness of mail ballots, making them a viable choice for increasing numbers of voters.

In reviewing EAVS data from the 2008 presidential election, Charles Stewart III of MIT suggested that as many as 7.6 million voters dropped out of the voting “pipeline” between requesting a mail ballot and having it counted, and called for increased attention to the process steps where those losses might be happening. That year, respondents from the two all-mail election states of Oregon and Washington reported below-average confidence that their votes were counted—and their confidence ratings corresponded with their support or opposition for mail-based elections.

As rates of voting by mail continue to increase, maintaining voter confidence is a critical concern in ensuring a healthy democracy. Election officials are increasingly offering ballot-tracking tools that allow voters to verify that their votes were received and counted.

Three states (Colorado, Oregon, and Washington) conduct elections primarily by mail, sending every voter a ballot at the start of the election cycle, while two others (California and Utah) allow counties to conduct all-mail elections as well. These states each provide tools where voters can look up the status of their ballot. Several also track ballots through the postal service, using USPS Intelligent Mail barcode scan data to provide more detailed

---

48 Id. at 598.
information about the ballot’s location. In addition to online lookup tools, some states also provide email or text messages with ballot tracking updates.

Ballot tracking technology provides multiple benefits: it offers election officials the opportunity to intervene and correct problems in the postal system, provides voters with clear confirmation that their ballots were counted, and generates data on ballot-mailing patterns that allow election officials to review and improve ballot-mailing procedures on an ongoing basis.

Another 22% of voters cast their ballots in-person before Election Day.51 Like mail voting, early voting is an increasingly popular choice, even more so in presidential election years. Though early voting has not been demonstrated to increase turnout, it can reduce the risk of Election Day issues by reducing stress on individual polling places.52

Like mail voting, early voting is increasingly popular with voters as more states offer the option. Though it is possible to administer early voting without technological support, tools like digital e-pollbooks that can sync voter turnout records with a central voter registration list in (or near to) real-time and on-demand ballot printing streamline the process in ways that have reduced the cost and complexity of setting up early voting centers for many states, and given voters additional convenience in deciding when and how to cast their ballots.

In these cases, technology offers election administrators resiliency in the face of changing voter needs. As public health considerations push many states to expand mail voting significantly in 2020, states with robust automation and well-maintained voter registration records are positioned to transition more smoothly. The systems they have built are less visible to the public, but have already had significant impact on rates of both registration and participation. In a moment of disruption, they continue to support election officials’ changing needs. Core election technologies are critical to supporting robust voter participation.

V. GUIDING PRINCIPLES

When discussing how technology can support voter turnout and civic engagement, “just add software” is a dangerous approach.53 Looking to

---

51 U.S. ELECTION ASSISTANCE COMM’N, supra note 30, at 11.
examples of successful uses of technology in inviting, informing, and modernizing voting, what common lessons do these cases offer?

When designing tools that support voters and the voting experience, three necessary (but not sufficient) requirements are accuracy, consistency, and accessibility. Where any of these principles are not taken into consideration, new tech runs risks of irrelevance, undermining civic trust, or even disenfranchising those it seeks to help.

Of these three principles, accuracy is both the most important and the clearest to evaluate. The principle of accuracy requires that the information offered by any tool or organization must first and foremost be correct. Democratic participation requires a wide-ranging dialogue, including discussion of information that may be controversial or in flux. As a subset of this conversation, information about how to vote is comparatively settled. Though an individual voter may not know their options or every step of the process, the answers to their questions are specific and verifiable.

However, because election processes vary so much by state (and in some cases, by county), accuracy can pose a significant challenge for any nationwide effort. As states offer additional means of registering and voting, those new options often add complexity as well. At least five states have different deadlines for registering to vote depending on whether the registration is submitted online or by mail. Some states set consistent early voting periods, while others allow local jurisdictions to set their own.

When done right, technology tools can help by tracking these state and local differences, geocoding voters, and ensuring that locally-accurate information gets matched to the people who need it. And yet, errors abound. In 2016, two major voter registration sites listed an incorrect phone number for the national election protection hotline. In 2018, Vox re-issued a registration deadline chart multiple times with corrections, even as outdated versions ricocheted around Twitter. Even official election websites occasionally list outdated, contradictory, or confusing instructions.

---

56 @govoteplz, TWITTER (Sept. 8, 2016, 2:09 PM), https://twitter.com/tianaej/status/773991749236686848 [https://perma.cc/6RT5-QLBA].
57 @voxdotcom, INSTAGRAM, https://www.instagram.com/p/BowgmkmgrQf/ [https://perma.cc/7QLQ-RNSN] (deadline graphic on Instagram included the caption “CORRECTION: A previous version of this graphic (it’s blue) contained incorrect or outdated information on the voter deadlines for Connecticut, Louisiana, and Utah. Florida and North
Where voters find accurate information, simply presented, they are more able to vote. Where design flaws or factual errors detract from this goal, technology can cause harm by distributing errors widely or contributing to voter confusion.

While accuracy ensures that a tool or organization serving voters is correct, the value of consistency recognizes that voter-facing messages must also work to reinforce any information that voters already have from other sources and to build confidence from that familiarity. Any technology used to invite or inform can increase trust through repetition or undermine it by contradicting (or appearing to contradict) other sources.

A potential voter setting out to learn about an election and deciding whether or not to take part will ask many questions. Especially for new voters, the path to casting a ballot is rarely straight. A voter on the Center for Civic Design’s sample “burdened path” lacks information about their options for voting or may discover some options too late to take advantage of them.\(^{58}\) They may worry about voter ID, even if their state does not require one. They move and need to update their voter registration just before the election. In this search for information, they may consult a wide variety of resources, from conversation to Internet search tools to social media engagement, to answer these questions.

In doing so, those voters will build up an understanding of the voting process and a corresponding expectation based on many separate inputs. Marsha Bates’s berry-picking model of search describes a discovery process in general terms: a person may begin with one question (“so, what is this election about?”) and, based on the first pieces of useful information found, begin to ask new questions that refine and build on the initial question (“how do these choices affect me? Do I care?”)\(^ {59}\)

Technologies for voter engagement can fall short of offering consistency in two ways: first, where they rely on jargon, and second, where they neglect to account for local variation. Official government materials can fall prey to jargon—long familiarity with these terms makes it difficult to remember that even terms like “early voting” are not immediately understandable to voters. One state offers voters requesting an absentee ballot the option to declare that:

Carolina have also adjusted their deadlines due to recent weather conditions. It has been corrected with confirmation from their state election offices. Thanks to our readers for helping point those out—and be sure to check with your local election office if you have questions, as dates are subject to change”\(^ {58}\).


“I am requesting a ballot for the presidential primary election and I may be absent on the day of the election from the city, town, or unincorporated place where I am domiciled, but the date of the election has not been announced. I understand that I may only make such a request 14 days after the filing period for candidates has closed, and that if I will not be absent on the date of the election I am not eligible to vote by absentee ballot.”

Even though there may be a legal requirement that mandates this language, this jargon makes it unlikely that any voter will choose this option.

Finally, any technology intended to support voter engagement and turnout must be accessible: it serves real people with a wide variety of needs. In this context, simplifying or omitting edge cases can exclude entire populations from participating in democracy. Where technologies simplify voting or provide a warm invitation only to groups who are already well-represented, they undermine the goal of robust participation.

Accessibility in a voting context includes a wide variety of considerations, from designing for low-vision and blind voters using screen-readers, to using clear language for voters with cognitive disabilities, to including support for multiple languages or publishing information on enfranchisement for returning citizens. Technology may need to address, or at least consider, transportation options for getting to the polls, how to procure voter IDs, or untangling residency requirements for students or people with unstable housing. Voters have many specific needs that affect their ability to take part.

Accessibility also means careful attention to design and presentation that can make the complexity of our election system into a process voters feel confident navigating. For example, the research and design phase of Los Angeles County’s “Voting Solutions for All People” project made “easy and accessible” one of the four guiding principles in the system’s approach to serving voters. The resulting prototypes include adjustable screens, multiple

---


61 In software development, these refer to less-common uses or user needs. Work to address edge cases is often less-prioritized than work that serves a larger number of anticipated users.


language support, and other accessibility features as core elements of the voting machine’s design.\footnote{Concepts, LOS ANGELES COUNTY REGISTRAR-RECORDER/COUNTY CLERK: VOTING SOLUTIONS FOR ALL PEOPLE, https://vsap.lavote.net/design-concepts-2/ [https://perma.cc/DWZ2-DY4F].} These voting machines are not yet in use in Los Angeles County, but they model how accessibility planning can support all voters.

In elections, good design goes a long way. No one technology may serve every one of those voters’ needs, but tools that do not meaningfully address any of them can suppress votes.

VI. CONCLUSION

This November, Election Day may hardly resemble the school-gymnasium scene many of us are accustomed to. In states like Colorado, it already does not. Voters register on state websites, seek out ballot and deadline information online, receive their ballots by mail, and track their return via text message. Where new technologies have allowed election administrators to offer voters greater choice and convenience, those same measures now support a robust election even in a time of social distancing.

How we take part in this election may feel new for many American voters, but it builds on existing technology. When Facebook and Snapchat ask their users to request mail ballots, and Google results include step-by-step instructions, they’ll be using existing systems to distribute new and changing information. If any states choose to mail every voter a ballot, that will be supported by their centralized databases and list-maintenance work through ERIC.

And if Election Day looks like putting on the “I voted” sticker that came with your ballot in the mail two weeks prior and publishing a selfie from the quiet of your home, you’ll simply be joining the future of voting that technology has long been building—just accelerated a bit.
AMPLIFY THE PARTY, SUPPRESS THE OPPOSITION: SOCIAL MEDIA, BOTS, AND ELECTORAL FRAUD

Samuel Woolley* & Nicholas Monaco**

CITE AS: 4 GEO. L. TECH. REV. 447 (2020)

TABLE OF CONTENTS

I. INTRODUCTION .................................................................................................................. 448
II. THE NEED FOR INFORMED POLICY ON POLITICAL BOTS ........................................ 451
III. CASE STUDIES OF POLITICAL BOT USAGE DURING ELECTIONS ...................... 451
      A. United States ............................................................................................................. 451
      B. North Macedonia ................................................................................................... 453
      C. Nigeria .................................................................................................................... 454
IV. ALTERNATIVE FORMS OF ELECTION-INFLUENCE BOTS: CRAWLER
      BOTS AND VOLUNTEER BOTNETS ........................................................................... 455
      A. Crawler Bots .......................................................................................................... 455
      B. Volunteer Botnets ................................................................................................... 456
V. TECHNOLOGICAL METHODS FOR CURBING BOT INFLUENCE ............................... 458
      A. Platform-Dependent Research ............................................................................. 458
      B. Cyborg Behavior .................................................................................................... 458
      C. Methodological Limitations and Asymmetric Visibility ....................................... 459
VI. THE FUTURE OF POLITICAL BOT MANIPULATION ................................................. 460
VII. CONCLUSION ............................................................................................................... 461

* Assistant Professor, School of Journalism, University of Texas at Austin. Program director, propaganda research, Center for Media Engagement at UT. Ph.D., University of Washington. I would like to sincerely thank the Knight Foundation, Omidyar Network, and Open Society Foundations for supporting my research; my colleagues in the Moody College of Communication at UT for their tireless help and guidance; and the team at Georgetown for convening the Election Integrity Symposium.

** Director, Digital Intelligence Lab at Institute for the Future. M.S., University of Washington. I am deeply grateful to: Georgetown Law School for convening the Election Integrity Symposium for which this article was written; Graphika, for hosting and facilitating research in this article; Daniel Arnaudo; Benecia Carmack, whose legal expertise helped inform the author's understanding herein; and the Georgetown Law Technology Review for their edits. Any mistakes are solely those of the authors.
I. INTRODUCTION

In the weeks preceding the 2016 United States presidential election several images began making the rounds on Twitter. The graphics sought to “remind” Democrats that they could vote via text message. According to The Wall Street Journal, they were built to resemble genuine “get-out-the-vote material produced by Hillary Clinton’s campaign.”¹ Many included the “Paid for by Hillary for President 2016” disclaimer that appeared in Clinton’s actual social media advertising. Some were in Spanish, targeting Latinx voters, while others included a photo of an African American woman holding an “African Americans for Hillary” poster. However, no states actually allow people to vote over text.

Many such vote-via-text tweets and corresponding images that circulated during the 2016 U.S. elections were spread by anonymous accounts—which lacked information, imagery, or content that identified the person behind them. Some accounts appeared to be spammers, with product advertisements interspersed among political messages. Many featured the hallmarks of social bots, or automated social media accounts built to look like real users and spread content. These “political bots,” so designated because most of their content was geared towards the manipulation of public opinion during a pivotal political event, worked to amplify the disinformative voting messages.² The logic behind these political bots is that if ten human-run accounts can spread many messages over Twitter, then one thousand automated profiles can spread masses.

Such examples of social media-based voter suppression, and others like them from around the globe, reveal the extent to which social media disinformation campaigns are tied to electoral fraud. Individuals and groups in a variety of countries have utilized these methods in attempts to both stymie the votes of the opposition and, often simultaneously, amplify information pertaining to their party, position, or cause. Suppression posts from these political actors may often implicitly and explicitly target minority communities with false information on elections and voting. According to a series of reports from the Institute for the Future’s Digital Intelligence Lab, “vulnerable demographics within already marginalized social groups—including young people, the elderly, military veterans, and those for whom English is a second language—were often particular targets of disinformation and harassment.” Moreover, “in already embattled minority and social groups, both human groups and bot armies spread harassing content and misleading information about voting and worked to foment in-group fighting

---


5 Id.
both online and offline.” The human consequences of these campaigns are, in other words, very much tied to adverse experiences among these marginalized groups.

Posts encouraging people to vote via text, or those directing people to vote on a day after an official election day, are a form of “fabricated content” which Wardle and Derakhshan describe as a type of “misinformation, disinformation, [or] mal-information.” These messages are spread using the strategies of computational propaganda, which constitutes “the assemblage of social media platforms, autonomous agents, and big data tasked with the manipulation of public opinion.” In the case of the false vote-via-text Clinton images from 2016, the images constituted disinformation: purposefully-spread false information. Disinformation online can often lead to the spread of misinformation: unintentionally-spread false content. As Jack argues, “The words we choose to describe media manipulation can lead to assumptions about how information spreads, who spreads it, and who receives it. These assumptions can shape what kinds of interventions or solutions seem desirable, appropriate, or even possible.” With this in mind, this paper is mindful of specific terms when describing media manipulation campaigns and moving into discussions of—and recommendations for—policy and the law.

There are significant legal implications for the use of political bots and other forms of computational propaganda during elections. This is particularly true when either automated or human-run accounts are leveraged for voter suppression specifically and electoral fraud more generally. Howard et al. opine that “election law in the United States, which already treads carefully in light of free speech issues, seems barely able to regulate political bots.” They build upon ideas of how political bots might run afoul of the law vis-à-vis campaign oversight, writing that “the conduct of political bots implicates some of the core issues of campaign regulations—including the ban on coordination between candidates and supporters, rules around soliciting financial support, and requirements to disclose affiliations.”

---


7 Woolley & Howard, supra note 2, at 4885–86.


10 Id.
II. The Need for Informed Policy on Political Bots

Globally, laws for regulating the use of political bots remain in the early stages of development. Examples of proposed and realized legislation in countries, including the United States and Brazil, have tended towards being overly broad and practically unenforceable. The state of California’s “bot bill”—SB 1001—aims to force certain types of bots (or, rather, their makers) to identify themselves online as automated.11 Wired magazine sums up the law as “noble, flashy, intriguing . . . and inept.”12

The issue with bills like SB 1001, and others like Senator Feinstein’s proposed “Bot Disclosure and Accountability Act,”13 is that they are overly broad in their definition of bots. Rather than focusing on politically-oriented accounts purporting to be real users, these efforts attempt to regulate all automated accounts. While the idea of identifying potentially manipulative automated accounts on social media platforms is laudable in a general sense, on a more granular level it overlooks the fact that bots play an infrastructural—and many times useful—role across social media and the Internet writ large.

There is a clear need for informed political bot policy. It is crucial, however, that the laws and policies created to deal with the issue of computational propaganda are written in collaboration with public interest technologists and political bot experts. Otherwise, such legislation runs the risk of censoring certain types of beneficial or benign communication. Put more simply, such legislation runs the risk of being technologically impossible or impractical to oversee and carry out.

III. Case Studies of Political Bot Usage During Elections

A. United States

Automation and other forms of coordinated, inorganic, manipulation campaigns have played a role in online political communication in the United States since at least 2010.14 In fact, early research on “astroturf” digital political campaigns suggests that the United States was the first country to experience “computational propaganda” during a major national election.15

---

11 SB 1001 was passed in 2018 and codified at CAL. BUS. & PROF. CODE § 17940–43.
15 Samuel Christopher Woolley, Manufacturing Consensus: Computational Propaganda and
Collections of political bots, organized human groups, and combinations of the two have been used in a number of attempts to game public opinion during a number of elections and major political events since then. Both the 2016 U.S. presidential election and 2018 midterm elections saw the use of coordinated, inorganic, “information operations” to sway, polarize, and disenchant voters across the country.\(^1\)

The case of the false Clinton “vote-via-text” messages on Twitter is one example of social media—and automated tools like political bots—being used to perpetuate electoral fraud and voter suppression in the U.S. In other instances, between 2010 to 2020, platforms like Facebook, Twitter, and YouTube have also been leveraged to intimidate voters and amplify unfounded claims of electoral cheating. According to VineSight, a firm working to fight social media misinformation, political bot accounts magnified misinformed claims of attempted vote manipulation during the 2019 Kentucky and Louisiana elections. These efforts corresponded with robocall campaigns urging voters to report alleged voter fraud.\(^2\)

According to ProPublica, these automated campaigns could foreshadow attempts to undermine the legitimacy of electoral outcomes in 2020 and onwards.\(^3\) In response to the 2019 efforts, University of Kentucky Law professor Joshua Douglas argued that “the way in which we handle these sorts of allegations from a losing candidate in 2019 will tell us if our democratic norms can sustain the same thing in 2020.”\(^4\)

Howard et al. add that political bots and other organized political campaign social media profiles can be put to another harmful use beyond

---


\(^4\) Id.
direct electoral fraud: evading or circumventing campaign finance regulation. Specifically, they suggest that political bots have the potential to muddy the already murky waters of campaign finance law by interfering with limitations on expenditures, limitations on contributions, and rules requiring disclosure. They argue that political bots can be, and have been, used to further illicit political campaign-Super PAC coordination and zombie electioneering (communication during campaigns that gives the appearance of large-scale support for an idea of political candidate through automated bots, comments, or other means.) Moreover, they build upon research by Indiana University, which found that political bots have been used to perpetuate smear campaigns against politicians—including, for example, Chris Coons during his bid for U.S. Senate in 2010—in efforts to discredit them and their campaigns.

B. North Macedonia

During the 2016 U.S. presidential election, Macedonia, now known as North Macedonia, earned itself a reputation for international disinformation after a small number of financially motivated individuals in the city of Veles spread false news stories. Two years later, the country was home to a coordinated, astroturfing campaign which attempted to decrease voter turnout in a domestic referendum.

In September 2018, Macedonia held a consultatory referendum on whether to formally change its name to the “Republic of North Macedonia.” The name change was a proposed compromise with Greece, which itself has a region named Macedonia within its borders and had long blocked the Balkan republic’s accession to the EU and NATO over the issue. The referendum was the first in a series of steps agreed upon by the North Macedonian and Greek governments in the Prespa Deal. Though the results of the referendum were non-binding, high turnout was crucial for ensuring that the government was undertaking the name change with the consent of the people.

In the months leading up to the referendum, a small number of hyperactive accounts appeared online calling for a boycott of the referendum under the hashtag #Бојкотирам (#boycott). Two aspects of the online

20 Howard, Woolley, & Calo, supra note 9, at 92.
21 See id. at 88.
campaign to boycott the referendum were noteworthy: (1) the stark coordination online, and (2) the role the Macedonian diaspora played in driving the campaign. Analyses from the Atlantic Council’s DFRLab and the network analysis firm, Graphika, both showed a high degree of automation and coordination among a small number of hyperactive accounts. The Transatlantic Commission on Election Integrity (TCEI) also noticed a large number of bots participating in the Twitter discussion around the referendum and noted that 10% of the accounts in their dataset had been created within 60 days of the referendum.\footnote{Press Release, Transatlantic Comm’n on Election Integrity, Alliance for Democracy, Macedonia Referendum: Twitter Bots Up Their Activity (Sept. 26, 2018), https://mailchi.mp/allianceofdemocracies.org/macedonia-referendum-twitter-bots-up-their-activity [https://perma.cc/9VLA-XLQ9].}

Interestingly, these analyses and subsequent reporting showed that much of the fabricated activity and energy online appeared to be coming from the Macedonian diaspora. For instance, one of the websites frequently promoted by this small group of users, bojkotiram.mk, was developed in part by a Macedonian living in Norway.\footnote{Kostas Zafeiropoulous, Alexander the Bot: The Twitter War for the Macedonian Soul, BALKININSIGHT (Dec. 18, 2019), https://balkaninsight.com/2019/12/18/alexander-the-bot-the-twitter-war-for-the-macedonian-soul/ [https://perma.cc/R8AF-ZKQK].} The fabricated traffic contained disinformation that opposed the referendum and sought to discredit the Prespa Deal. Although voters overwhelmingly approved the name change, turnout for the referendum ended up being far under the 50% the government had hoped for, with only 37% of voters casting ballots.\footnote{Id.}

C. Nigeria

Nigeria also experienced similar fabricated content activities in 2019, although in a different context. During Nigeria’s national elections, a highly active botnet on Twitter promoted messages questioning the legitimacy of the electoral process and encouraging citizens not to vote. This botnet also promoted Biafra, a region in Eastern Nigeria whose attempted secession in the 1970s precipitated a civil war in Nigeria. The botnet-promoted messages alleged that Nigerian democracy was a sham, and called on voters to boycott the election. While relatively unsophisticated, this botnet produced a high volume of tweets encouraging citizens to “stay home” instead of voting.\footnote{Nigeria’s General Elections: The Online Information & Media Landscape, CTR. DEMOCRACY AND DEV. (Feb. 20, 2019), http://www.cddwestafrica.org/nigerias-general-elections-the-online-information-media-landscape/ [https://perma.cc/323T-MTFD].}
Outside the electoral process, automated agents have placed other democratic processes into its crosshairs. In the realm of free speech, a large botnet of over 500 accounts regularly co-opts trending topics on Twitter to promote websites alleging the innocence of a religious leader in India, Rampal Singh. Singh has been behind bars since 2014 for multiple charges of murder that occurred in his Ashram. Most recently, this same large botnet co-opted the October 2019 U.S. Democratic debates’ hashtag to push their messaging.

IV. ALTERNATIVE FORMS OF ELECTION-INFLUENCE BOTS: CRAWLER BOTS AND VOLUNTEER BOTNETS

Much of the reporting and research on political bots concentrates on megaphone bots: bots that post or promote messages at scale with the intent to spread a specific message. It is, however, important to note that more passive forms of political bots exist and can be used to influence elections.

A. Crawler Bots

One form of these political bots are crawlers (or “scrapers”). Crawlers are programs that visit websites, including social media sites, to gather data. As the cybersecurity company Imperva has reported, these bots are extremely common on the Internet and often account for more online traffic than humans. In a political context, crawlers can be used to gather information on target voters and demographics online—for instance, to gauge public reaction to political events and inform real-time campaign and messaging strategy. In Taiwan, as early as the 2014 Taipei mayoral election, crawlers were used to gather data on 11–14 million users—roughly half of the island’s population.

---

Data collected from these bots can also be used to enhance microtargeting efforts aimed at increasing or suppressing voter turnout. This usage is all the more pernicious given the fact that de-anonymization is easier when combining multiple datasets. Enhanced “persuasion profiles” can be compiled on specific targets to aim for desired outcomes—such as opting not to vote. Indeed, this aim was part of the strategy used by Cambridge Analytica in the 2016 U.S. presidential election.

A final note on crawler bots is that poor Internet policy is likely to exacerbate their abuse in the United States. In particular, the 2017 adoption of Senate Joint Resolution 34, which prevented the Federal Communications Commission from limiting Internet service providers’ ability to harvest and sell consumer data, opens the floodgates for new possibilities of manipulative microtargeting. The potential for harm abounds: one compelling example is Emerdata, the company formed by former employees of Bell Pottinger and Cambridge Analytica in the wake of the latter’s demise. Emerdata has been explicit about its intent to use all publicly procurable data, including search histories and information purchased from data brokers, to microtarget voters in increasingly sophisticated ways in countries around the world. Senate Joint Resolution 34 is tantamount to an invitation for Emerdata and others to do their worst.

**B. Volunteer Botnets**

Perhaps the most surprising new form of automated agent to arrive in recent years is the volunteer botnet. The idea behind volunteer botnets is simple: users consent to allow political campaigns to use their accounts to post messages when desired. The Russian Ministry of Foreign Affairs used this technique to promote its ambassador to the United Kingdom in 2017, but this...
tactic has increasingly appeared within an electoral context in past years to promote candidates and parties. In 2018, Italy’s far-right populist party La Lega Nord used volunteer botnets to promote its party and current Prime Minister Matteo Salvini. The same year, Mexico’s current president Andrés Manuel López Obrador’s party MORENA allowed users to volunteer their Twitter accounts to spread automated messaging supporting their party at treinta.moren.org during the presidential election.

In perhaps the most fascinating case so far, young Labour activists organized in 2017 in the United Kingdom to build a Tinder volunteer botnet. Users “donated” their Tinder profiles to a computer for two hours at a time, during which time the bot would use a set of predefined criteria to attempt to persuade voters between 18–25 years of age to vote for the Labour party in upcoming parliamentary elections. The team behind the bot estimated it sent between 30,000–40,000 messages to 18–25 year olds in marginal constituencies, with the explicit goal of “oust[ing] the conservative government.” In the end, the age group targeted by the Tinder bot displayed its highest election turnout since 1992.

The UK Tinder bot’s developers were overt about their tactics, and even raised funds to build the bot publicly on Indiegogo. While several media outlets praised the political savvy of the strategy, Oxford Internet Institute researchers Robert Gorwa and Doug Guilbeault were quick to point out the bots may not have been praised had they been promoting the conservative Tory party. Gorwa and Guilbeault also rightly highlight that the strategy contains serious ethical concerns—chief among them being that users were not aware that they were interacting with a bot. This subterfuge is the most

37 Silverman & Alexander, supra note 24.
40 Rasmussen, supra note 39.
insidious aspect of volunteer botnets: users can undertake normal activity—e.g., posting on Twitter or swiping on Tinder—while the bot simultaneously carries out its work. This mix of human and automated activity is known as “cyborg behavior,” and makes discerning human and automated behavior nearly impossible. Moreover, it is particularly challenging for companies and researchers to detect for reasons we explore below.

V. TECHNOLOGICAL METHODS FOR CURBING BOT INFLUENCE

A. Platform-Dependent Research

While tools for detecting bots on social media exist and have benefitted from widespread use in recent years, several notable limitations still exist. The most important limitation is that the most sophisticated detection tools are exclusively limited to Twitter, where rich metadata on users and an especially active social scientific research community have resulted in several machine learning tools that classify accounts as bots or humans. These include Indiana University’s Botometer, Bot Sentinel, Botcheck.me, and rweetbotornot.42 While bots are certainly active on other social media platforms such as Facebook, YouTube, Gab, Instagram, and even regular websites, detection is much harder, since granular data on users is unavailable unless one owns the servers on which the traffic occurs (i.e., the platform itself, in most cases). Unlike Twitter, external researchers on these platforms must exclusively deduce bot activity through manual investigation—signs such as “superhuman” activity statistics or large numbers of identical posts can help reveal an account’s automation. (For example, the authors have used these methods to detect bot activity on Gab.43) For these reasons, our understanding of bot activity on these platforms is necessarily quite limited. Without the development of tools for these platforms or more transparency from the companies themselves, our understanding of social bots’ activities outside of Twitter will remain limited.

B. Cyborg Behavior

“Cyborgs”—accounts that are only automated part of the time—present another challenge for bot detection tools. Most bot detection tools

42 As of publication, these websites can be visited, respectively, at: https://botometer.iuni.iu.edu/#/, botsentinal.com, botcheck.me, and https://github.com/mkearney/tweetbotornot.
currently available take an account as input, examine features in the account’s metadata (such as date of creation, whether the account has a default profile and banner photo, average posts per day, etc.) and output a “bot probability,” a percentage between 0–100% representing the probability that the input account is a bot. Researchers and academics have adopted the common convention of considering accounts with a higher than 50% probability to be automated.44

Cyborgs have become an increasingly frequent phenomenon on social media, especially since their ability to fly under the radar and evade traditional bot detection techniques is higher than a normal bot account—after all, they are operated by a human at times. The authors’ team at the DigIntel Lab recently detected cyborg activity promoting then-U.S. Democratic presidential candidate Andrew Yang during the September 2019 Democratic primary debates. Cyborgs also spread anti-vaccination disinformation during the following month’s debate.45

A partial solution to this problem is to search for “coordinated” behaviors on social media, rather than ascertaining whether isolated accounts are bots. To this end, researchers such as Ben Nimmo have iterated new techniques such as the Coefficient of Traffic Manipulation (CTM) for interrogating whether online behavior (such as hashtag promotion) is organic or coordinated.46

C. Methodological Limitations and Asymmetric Visibility

A final limitation of modern bot detection is methodological. Current detection tools draw on supervised machine learning techniques. Such techniques let algorithms learn which characteristics of an account are most predictive for determining whether an account is a bot or human. The first step in “teaching” such an algorithm is to assemble a training set: a set of bot accounts and a set of human accounts from which the algorithm learns its

associations.\textsuperscript{47} This stage also faces an epistemological limitation, which is rarely acknowledged in the research world—namely, the assumption that humans are able to accurately and reliably discern bots from humans online. This limitation is likely to exist for the foreseeable future because of a common problem in online social science research: asymmetric visibility. Asymmetric visibility refers to the fact that social media companies can see more data and information on their platforms than external researchers, and it is one of the predominant factors limiting understanding of political automated activity on social media at large.

VI. THE FUTURE OF POLITICAL BOT MANIPULATION

The future of political bot-based manipulation, and computational propaganda during elections more broadly, appears to be focused on campaigns aimed at minority populations, diaspora communities, and other marginalized groups. While individualized political communication is on the rise, demographic political targeting continues to play out online. Far from Cambridge Analytica’s claims that demographic politics is dead, it appears that the powers that be remain focused on manipulating voters on the outside, fringes or middle of the political spectrum—even when political advertisers work to concentrate their messages on subsections, or even individuals, within a particular group.\textsuperscript{48}

It is also clear that diaspora populations are likely to increasingly function as propagators or targets of automated activity in the future. India attempted to mobilize Indian-heritage UK citizens to vote against the Labour party in 2019 UK parliamentary elections, in part because of their stance on Kashmir.\textsuperscript{49} Above, we highlighted that the Macedonian diaspora played an important role in funding both the boycott campaign and building bots and websites that were centrally involved in that campaign on Twitter. WeChat, a Chinese app used heavily by overseas Chinese for most online needs from social media to mobile payments, is also increasingly playing a role in political


\textsuperscript{48} See generally, Woolley, supra note 16.

campaigning in the United States and Canada. For instance, the Vancouver Friendship Society, an organization with ties to the Chinese government, is under investigation for an alleged WeChat vote-buying scandal in Canada’s 2018 parliamentary election.

VII. CONCLUSION

Political bots now play a crucial and manipulative role in both amplifying and suppressing online political content during elections. On sites like Twitter, Facebook and Reddit, these automated tools scale underhanded efforts to magnify particular perspectives and candidates. Because they are computationally enhanced, political bots are also especially useful in spreading disinformation to voters. These voter suppression and intimidation efforts, well-documented during the 2016 U.S. elections, have been used to target minority communities and issue-focused voters.

The use of bots in digital politicking is not confined to the U.S, however. This paper offers case studies that highlight similar automated electioneering campaigns in North Macedonia and Nigeria. The types of bots used during these information operations vary, and we have endeavored to describe them above. Looking forward, we expect bots to continue to play an important role in driving political communication—including disinformation and politically-motivated harassment—during elections. There are signs that automated “false news” campaigns are already contributing to online confusion and offline violence during important political events around the world. Moving forward, it is crucial that policy makers collaborate with public-interest technologists to generate sensible and technically feasible laws and regulations to curb the influence and impact of political bot campaigns.


STUDENT NOTE: THE PERNICIOUS PROBLEM OF PLATFORM-ENABLED VOTER INTIMIDATION

Christopher Conrad*

CITE AS: 4 GEO. L. TECH. REV. 463 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 463
II. THE COLLISION OF VOTER INTIMIDATION AND INTERNET PLATFORMS ................................................................. 465
III. THE LIMITED LANDSCAPE OF VOTER INTIMIDATION LAWS PRIOR TO 1965 ....................................................................................................... 472
IV. SECTION 11(B) OF THE VOTING RIGHTS ACT & PLATFORM-BASED VOTER INTIMIDATION ........................................................................ 475
   A. Section 11(b)’s Statutory Reach to Platform-Enabled Voter Intimidation ...................................................................................... 476
   B. Overcoming Constitutional Limitations .................................................. 485
   C. Obtaining Adequate Relief Under Section 11(b) ................................... 488
V. SOLUTIONS BEYOND EXISTING LAW ........................................ 490
VI. CONCLUSION .............................................................................. 497
APPENDIX ......................................................................................... 499

I. INTRODUCTION

From Reconstruction through the Civil Rights Era, voter intimidation typically took the form of flagrant, violent targeting of blacks and their

* Georgetown Law, J.D. 2020; The Elliott School of International Affairs, The George Washington University, M.A. 2015; Temple University, B.A. 2013. © 2020, Chris Conrad. For helpful comments and conversations, thanks to Julie Cohen, Lyn Abdullah, Josh Banker, and the participants of the Election Integrity Symposium. Thanks also to the hard-working staff on the Georgetown Law Technology Review, including Harsimardh Dhanoa and Laura Ahmed. Finally, I would like to express my deepest gratitude to my wife, Kate Jinghua Zhou, for her unwavering love and support.
political supporters. However, in recent decades, those attempting to suppress the vote have embraced more subtle, cynical, and creative methods. Increasingly, voters are being subjected to false information about voter requirements, aggressive questioning about citizenship, and anonymous threats of harm designed to deter them from voting.

As part of this transition to more inconspicuous forms of voter intimidation, individuals and political organizations have largely supplanted local law-enforcement officials and white-supremacist groups as the main perpetrators. Instead of polling places, these actors have taken to Internet platforms to suppress the franchise of minority voters. And for good reason: platforms are optimally suited for voter intimidation. Internet platforms allow for anonymous speech, amplify and polarize narratives, and can be manipulated through use of bots. The opaque environments these platforms form make it difficult for voters to confidently calculate risks to their safety and freedom when crowd-sourced threats and deception loom large.

Worse yet, voter intimidation on platforms often cannot be identified and removed quickly enough to offset its impact on voter behavior. Remedies available under current law—ex post relief limited to injunctions and complicated by burdensome mens rea requirements—generally do little to deter platform-enabled voter intimidation that occurs contemporaneously with elections. Consequently, eliminating this form of voter intimidation has thus far fallen on the platforms themselves, which appear ill-equipped and, sometimes, disincentivized to privately address the problem.

1 See Rayford W. Logan, The Betrayal of the Negro: From Rutherford B. Hayes to Woodrow Wilson 57 (1954) (“[T]he fear of potential, and in some instances actual, intimidation prevented most Negroes from voting in the South . . . .”); see also id. at 378 (“Negroes, moreover, suffered open or concealed intimidation when they tried to vote.”).


3 Voter Intimidation, ACLU, https://www.aclu.org/sites/default/files/field_pdf_file/kyr-voterintimidation-v03.pdf [https://perma.cc/ZQ87-R5SJ]; see also Cady & Glazer, supra note 2, at 215.


5 Phillip N. Howard, How Political Campaigns Weaponize Social Media Bots, IEEE SPECTRUM (Oct. 18, 2018, 3:00 PM), https://spectrum.ieee.org/computing/software/how-political-campaigns-weaponize-social-media-bots [https://perma.cc/5VJJ-EMB5] (“Bots’ are just bits of software used to automate and scale up repetitive processes, such as following, linking, replying, and tagging on social media. [T]hey can . . . affect public discourse by pushing content from extremist, conspiratorial, or sensationalist sources, or by pumping out thousands of pro-candidate or anti-opponent tweets a day. These automated actions can give the false impression of a groundswell of support, muddy public debate, or overwhelm the opponent’s own messages.”).
This Note analyzes the challenges of deterring platform-based voter intimidation in search of viable alternatives. Part I of this Note provides a brief overview of voter intimidation in the United States and discusses its recent collision with social media and other third-party Internet forums. It details how the unique characteristics of platforms enable voter intimidation. It also provides illustrations from the last several federal election cycles. Part II surveys the landscape of criminal and civil voter-intimidation laws that were adopted prior to the Voting Rights Act of 1965. It finds that none of these laws are adequate to deter platform-enabled voter intimidation. Part III then examines Section 11(b) of the Voting Rights Act of 1965 (“VRA”), analyzing whether it can stem the rising tide of platform-enabled voter intimidation. Though Section 11(b)’s provisions are effective in a few circumstances, the Part concludes that the VRA is ill-suited to address most forms of platform-enabled voter intimidation. In turn, Part IV explores other viable remedies. It narrows the field of options to a Hobson’s choice between the status quo and the implementation of a costly multi-day election format at the national level which may or may not effectively mitigate the effects of same-day platform-enabled voter intimidation. This Note ultimately concludes that practical and constitutional limitations compel adoption of a “content takedown” response to voter intimidation that centers on collaboration between government regulators and large private platforms.

II. **The Collision of Voter Intimidation and Internet Platforms**

The rise of Internet platforms has substantially altered the form and substance of voter intimidation. During and prior to the Civil Rights Era, those seeking to undermine suffrage deployed physical violence as their instrument of choice. Mobs beat and murdered blacks and their political supporters. Perpetrators also targeted victims with economic coercion, malicious prosecution, and public identification—a historic analog to doxing. In Tennessee, landowners, merchants, banks, and corporations retaliated against black voters by refusing to transact with them for goods and services. In Florida, law enforcement officials guarded ballot boxes and arrested black

---

6 Third-party Internet forums include “message board and group blogs.” DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 143 (2014).
7 See CONG. GLOBE, 42d Cong., 1st Sess. 197 (1871) (statement of Sen. Ames) (noting that 859 political murders of Republicans had occurred in Louisiana in recent years, none of which were prosecuted); see also Paynes v. Lee, 377 F.2d 61, 63 (5th Cir. 1967); United States v. Wood, 295 F.2d 772, 776 (5th Cir. 1961); Michael Kent Curtis, The Klan, the Congress, and the Court: Congressional Enforcement of the Fourteenth and Fifteenth Amendments & the State Action Syllogism, a Brief Historical Overview, 11 U. PA. J. CONST. L. 1381, 1399 (2009).
In Louisiana, the Ku Klux Klan (“KKK”) generated and distributed handbills to identify the names and addresses of blacks who had registered to vote. These methods mostly met their overdue demise in the 1960s and 1970s, snuffed out through coordinated efforts of the federal legislative, judicial, and executive branches.

But voter intimidation lived on, undergoing a pernicious evolution. In place of in-person beatings and boycotts, arm’s length tactics took over. During the 1993 New York City mayoral race, signs were anonymously posted in Hispanic neighborhoods falsely reporting that federal immigration authorities would be monitoring polling places. In 1998, a South Carolina state representative mailed thousands of notices to individuals in predominantly black neighborhoods that warned voters “this election is not worth going to jail”—a familiar threat steeped in a long history of voter intimidation in the South. And in 2006, a mass-mailing service, acting under the direction of a candidate for U.S. Congress, sent out 14,000 letters falsely informing registered voters with Hispanic surnames that if they voted, their


10 See United States ex rel. Katzenbach v. Original Knights of KKK, 250 F. Supp. 330, 342 (E.D. La. 1965) (describing intimidating KKK posters designed to identify specific individuals and businesses that the KKK was targeting); see also Voting Rights: Hearings Before the S. Subcomm. on the Judiciary, 89th Cong. 1292 (1965) (detailing a systematic campaign of intimidation in Haywood County, Tennessee in 1960, after more than one hundred whites developed and circulated a list of black citizens to be denied voting credit); Brief for the Campaign Legal Center as Amicus Curiae Supporting Plaintiffs at 21, League of United Latin Am. Citizens - Richmond Region Council 4614 v. Pub. Interest Legal Found., 2018 WL 3848404 (E.D. Va. Aug. 13, 2018) (No. 1:18-CV-00423) [hereinafter CLC Amicus Brief].


12 See President Lyndon Johnson, Address to Congress on Voting Rights (Mar. 15, 1965) (“At times, history and fate meet at a single time in a single place to shape a turning point in man’s unending search for freedom. So it was at Lexington and Concord. So it was a century ago at Appomattox. So it was last week in Selma, Alabama. There, long-suffering men and women peacefully protested the denial of their rights as Americans.”); see PEOPLE FOR THE AM. WAY FOUND. & NAACP, supra note 11, at 3 (describing the coordinated passage of the 1965 Voting Rights Act as “among the crowning achievements of the civil rights era, and a defining moment for social justice and equality”).

13 PEOPLE FOR THE AM. WAY FOUND. & NAACP, supra note 11, at 10.

personal information would be collected by a government computer system that could be used to target illegal immigration.\textsuperscript{15}

As these incidents have proliferated, so too has the use of Internet platforms.\textsuperscript{16} These platforms—which include social media sites and a fragmented landscape of third-party Internet forums—form vast universes of connected users.\textsuperscript{17} Their ubiquity endows them with unprecedented power to shape and control access to information, and their rise has fundamentally altered the process by which individuals communicate with one another.\textsuperscript{18} The Supreme Court has likened these platforms to “the modern public square” and recognized that they allow individuals to “explor[e] the vast realms of human thought and knowledge” and to “engag[e] in the legitimate exercise of First Amendment Rights” by accessing “the world of ideas.”\textsuperscript{19}

However, platforms also present risks that threaten to undermine suffrage. For one, they facilitate anonymous speech. When users believe that their acts will not be attributable to them, they are more likely to defy social norms and act destructively without fear of external sanction.\textsuperscript{20} In traditional legal contexts, the peril of anonymity has long served as a targeted evil for criminal and civil legislation. Many states, for example, have enacted anti-mask statutes to prevent crimes “daringly committed, under the protection of masks and other disguises.”\textsuperscript{21} At the federal level, the Ku Klux Klan Enforcement Act of 1871 (“the KKK Act”), a civil voter-intimidation statute, prefaces its ban on interference with the “support or advocacy . . . of the

\textsuperscript{15} United States v. Tan Duc Nguyen, 673 F.3d 1259, 1261 (9th Cir. 2012).


\textsuperscript{17} Howard Shelanski, Information, Innovation, and Competition Policy for the Internet, 161 U. PA. L. REV. 1663, 1682 (2013). Internet platforms have been described in varying terms. \textit{Compare} id. at 1665 (“digital platforms [are] products or services through which end users and a wide variety of complementary products, services, or information (‘applications’) can interact”) with Laurence Meyer, \textit{Digital Platforms: Definition and Strategic Value}, COMM. & STRATEGIES, 2d Quarter 2000, at 127, 128 (2000) (defining a digital platform as an “intermediation activity linked with the ‘assembly’ of content and services onto a coherent technical and commercial access platform”).


\textsuperscript{20} CITRON, \textit{supra} note 6, at 57.

\textsuperscript{21} The Second Circuit used this language to describe New York’s anti-mask law. See Church of Am. Knights of the KKK v. Kerik, 356 F.3d 197, 205 (2d Cir. 2004); see also id. at 203 (“New York’s anti-mask law, reenacted in its current form in 1965, can be traced back in substance to legislation enacted in 1845 to thwart armed insurrections by Hudson Valley tenant farmers who used disguises to attack law enforcement officers.”).
election of any lawfully qualified person . . .” with language contemplating the role of anonymity in fostering violence against minorities and their supporters.22

It should come as no surprise, then, that the anonymity facilitated by online platforms has proved an effective tool to exploit voters’ fear and to dissuade them from exercising their rights.23 In 2016, anonymous posters on the popular third-party forum, 4chan, coordinated plans to call Spanish-language radio stations and report that U.S. Immigration and Customs Enforcement (ICE) agents would be present at polling places.24 Two years later, Twitter suspended thousands of anonymous accounts that it suspected were involved in election disinformation campaigns coordinated by foreign governments.25

The manipulative power of online platforms is not only fueled by the ability of users to mask their identities. Platforms also enable ideologically homogenized individuals and groups to connect and organize to facilitate mass-intimidation. Historically, these individuals and groups had to physically venture out into the community in search of those sympathetic to their causes. With the emergence of digital communications platforms, gone are the physical and time constraints on human organization.26 As professor Danielle Citron observes, networked communications have removed “practical barriers that once protected society from the creation of antisocial groups.”27 As a result, the number of hate organizations has more than doubled over the last two decades.28

__22__ 42 U.S.C. § 1985(3) (2018) (“If two or more persons in any State or Territory conspire or go in disguise on the highway or on the premises of another . . .”).

__23__ See Cady & Glazer, _supra_ note 2, at 218.

__24__ Craig Silverman, _We’re Tracking All the Election Day Rumors, Hoaxes, And Debunks_, _BuzzFeed_ (Nov. 8, 2016, 11:19 PM), https://www.buzzfeednews.com/article/craigsilverman/all-the-election-debunks [https://perma.cc/3EAZ-7VPQ]; _see also_ Appendix 3.3.


__26__ CITRON, _supra_ note 6, at 62–63.

__27__ Id.

This dynamic also empowers political organizations, which are often responsible for mass voter intimidation. In September 2016, the Public Interest Legal Foundation (“PILF”), a law firm “dedicated . . . to election integrity,” published a report titled Alien Invasion via Twitter and Facebook. This report “characterize[d] the existing voter registration system as an ineffective honor system” and offered “the names, home addresses, telephone numbers, and, in some cases, social security numbers” of 1,852 “illegal registrants” registered to vote in Virginia. The plaintiffs in the case alleged that several of the “suspected aliens” listed in the report were not “aliens” at all; they were citizens who had been removed from the voter rolls for administrative reasons. Nonetheless, PILF, allegedly with full knowledge of these inaccuracies, disseminated the report via Facebook and Twitter, spreading it to politically-aligned news outlets that then shared it with ideologically-aligned readers.

In addition to demonstrating the organizational properties of platforms, PILF’s use of Facebook and Twitter also indicates how social media and third-
party forums can amplify and polarize speech. As danah boyd of Data & Society observes, tools originally built for “online community, communication, and information access” have been “weaponized to radicalize people toward extremism, gaslight publics, [and] serve as vehicles of cruel harassment.”

Online groups hear reverberations of their own voices when they are homogenized, and engage in cruelty competitions. Although viewpoint polarization originates in narrow online contexts, the Internet can prolifically spread abuse. Extreme posts draw attention and go “viral” in part because filtering algorithms used by major platforms highlight popular information. This process “lay[s] the groundwork for continued distribution” and contributes to what Citron terms “information cascades.”

Amplification and polarization enable speakers to stoke fear and hinder the free exercise of voting rights. During the 2016 presidential election, Facebook users made veiled and direct threats of violence targeting Clinton supporters. Their comments included “it’s LOCK AND LOAD if Donald J. Trump isn’t president,” and “I pray Trump wins as we veterans and military are ready to take our country back. You civilians either follow us or get out of the way. Your only other option is to be a qualified target.” These menacing posts, directed to connections within the speakers’ social networks, were visible to the broader public because of the viewing permissions set on the users’ accounts. As a result, they were picked up by third parties and reshared.

Unsurprisingly, anonymized messages, amplified by extremist communities for their vitriolic value, were even more polarized. Users in 4chan’s far-right-wing discussion board “/pol/” actively promoted the physical coercion of voters, posting comments like “already intimidating voters at stations [...] My SKS [semi-automatic rifle] is ready for whoever starts the process” and “What if we lose? Then we’ll meet IRL [in real life],” the latter comment appending a photo of grenades.

---

35 See generally PUB. INTEREST LEGAL FOUND., ALIEN INVASION II (May 2017), https://publicinterestlegal.org/files/Alien-Invasion-II-FINAL.pdf [perma.cc/RUA9-LW5B].
37 CITRON, supra note 6, at 63 (citing CASS R. SUNSTEIN, REPUBLIC.COM 2.0 60 (2007)).
38 CITRON, supra note 6, at 66–67.
39 Id.
40 Id.
42 Id.; see also Appendix 1.2 and 1.5.
43 See Appendix 3.1 and 3.2.
the broader public, this messaging forum—a hotbed of polarized political hate—bled into the mainstream during the 2016 election when bloggers and outside news outlets republished the incendiary comments.

Finally, harnessing the attributes of anonymity, organization, polarization, and amplitude, malicious actors have deployed platform bots—software that creates content and interacts with people—over recent election cycles to manipulate voter behavior.\(^{44}\) During the 2016 presidential election, a well-known pro-Trump Twitter bot, using the handle “@NeilTurner,” shared a doctored hoax image of an undocumented immigrant being arrested by an ICE agent at a polling station.\(^{45}\) The tweet, which was sent out to approximately 30,000 followers, spread rapidly throughout the Internet.\(^{46}\) Before the post was removed by Twitter, the photo was viewed and shared thousands of times, and was circulated widely among prominent news outlets.\(^{47}\)

As researchers of computational propaganda have observed,\(^{48}\) platforms are particularly susceptible to these types of incidents because they operate without editors who would maintain quality or control the circulation


\(^{48}\) The Computational Propaganda Project at the University of Oxford describes computational propaganda as the functional use of algorithms and automation to “manipulate public opinion by amplifying or repressing political content, disinformation, hate speech, and junk news.” See *The Computational Propaganda Project*, OXFORD INTERNET INST., https://comprop.oii.ox.ac.uk/ [https://perma.cc/BCA6-B9C3].
of content. Compounding the problem, bots are cheap and easy to use. Their operation and design also shift rapidly, allowing them to spread information while avoiding detection and removal. These tools, along with the platform characteristics they channel, make social media sites and third-party forums the perfect avenues for voter intimidation and manipulation.

III. THE LIMITED LANDSCAPE OF VOTER INTIMIDATION LAWS PRIOR TO 1965

In stark contrast to the alarming rise of platform-enabled voter intimidation, there are few legal remedies to protect suffrage from disruptive speech.

Criminal laws, in general, are ill-suited for prevention of modern voter intimidation. Enforced by the U.S. Attorneys and the Department of Justice, these laws require a robust showing of a defendant’s intent to intimidate, the existence of a conspiracy, or both. Because voter intimidation is often subtle and without witness, these elements deter prosecutions. Aside from statutory requirements, practical considerations also limit the effectiveness of criminal laws in combating voter intimidation. Prosecutors must elicit public testimony from victimized voters in adversarial proceedings—a daunting challenge for prosecutions of anonymous intimidation directed at broad, faceless audiences.

49 See Howard, supra note 5.
50 See Bob Abeshouse, Troll Factories, Bots and Fake News: Inside the Wild West of Social Media, AL JAZEERA (Feb. 8, 2018), https://www.aljazeera.com/blogs/americas/2018/02/troll-factories-bots-fake-news-wild-west-social-media-180207061815575.html [https://perma.cc/7LM7-2YCF] (“Seth Turin’s software is one of many turn-key systems to pop up in recent years that make it easy to build and use bots on social media sites. A version of his platform can be purchased for as little as $300.”); Zack Whittaker, Bots Are Cheap and Effective. One Startup Trolls Them Into Going Away, TECHCRUNCH (Feb. 5, 2019, 10:10 AM), https://techcrunch.com/2019/02/05/kasada-bots/ [https://perma.cc/U3MC-55XU].
51 See Marwick & Lewis, supra note 44, at 38.
53 See, e.g., 18 U.S.C. § 241 (2012) (making it a felony for two or more persons to “conspire to injure, oppress, threaten, or intimidate” any person in the free exercise of any right or privilege secured by the Constitution); 18 U.S.C. § 594 (2012) (prohibiting the intimidation, threat, or coercion of voters “for the purpose of interfering” with their right to vote) (emphasis added); 52 U.S.C. § 20511(1) (2012) (prohibiting anyone from “knowingly and willfully” intimidating or coercing an individual voting or attempting to vote).
54 PUB. INTEGRITY SECTION, supra note 52.
55 Id.
Prosecutors also face inherent resource limitations that make it difficult to justify prosecuting subtle online voter intimidation that has an imperceptible impact on an election that has already passed.56 Consequently, civil voter-intimidation statutes, which can be used by individuals, voting rights advocates, and political entities, offer a less constrained legal instrument to root out unlawful voter intimidation.57 These civil laws first surfaced in the wake of the Civil War as a response to the Ku Klux Klan’s widespread campaign of political violence against Southern blacks and their white supporters.58 Fearing a collapse of Reconstructionist policies, Northern Republicans passed the KKK Act. This law, designed to reach both state and private interference with the right to vote, provided:

If two or more persons in any State or Territory conspire or go in disguise on the highway or on the premises of another . . . to prevent by force, intimidation, or threat, any citizen who is lawfully entitled to vote, from giving his support or advocacy in a legal manner, toward or in favor of the election of any lawfully qualified person as an elector for President or Vice President, or as a Member of Congress of the United States . . . the party so injured or deprived may have an action for the recovery of damages occasioned by such injury or deprivation, against any one or more of the conspirators.59

The “Support and Advocacy” clause of the KKK Act is still in effect today. However, several Supreme Court decisions in the aftermath of its enactment have curtailed its use by private litigants.60 Its conspiracy element requires

57 See Cady & Glazer, supra note 2, at 179.
58 See Curtis, supra note 7, at 1399–1400.
59 42 U.S.C. § 1985(3) (2012); see also Curtis, supra note 7, at 1399.
60 In 1882, the Supreme Court held that neither the Fourteenth Amendment nor the Privileges and Immunities Clause of Article IV furnished Congress with the authority to regulate actions by private persons. United States v. Harris, 106 U.S. 629, 640, 643 (1882). The Court’s decision in Harris “chilled civil claims under [the civil provision of the KKK Act], which was assumed to have the same constitutional inadequacies as the corresponding criminal provision.” Catherine E. Smith, (Un)masking Race-Based Intracorporate Conspiracies Under
“some showing of the defendant’s mindset”—a formidable obstacle to successful rights vindication.61 And, while some scholars have argued that the KKK Act lacks an intent requirement, courts have found otherwise, placing additional burdens on private litigants at the pleadings stage.62

In the 1950s, with the KKK Act in a dormant state and the civil rights era taking hold, Congress furnished black voters with an additional civil remedy to combat interference with voting behavior. It passed the Civil Rights Act of 1957, which sought to curtail “public and private interference with the right to vote on racial grounds.”63 Section 131(b) of the Act reads:

No person, whether acting under color of law or otherwise, shall intimidate, threaten, coerce, or attempt to intimidate, threaten, or coerce any other person for the purpose of interfering with the right of such other person to vote . . . .

This statutory language improved on the KKK Act by omitting a conspiracy requirement. However, its text expressly required that plaintiffs demonstrate intent on the part of each individual defendant to act “for the purpose of” intimidating voters. Following Section 131(b)’s enactment, this element proved an extremely onerous burden on plaintiffs litigating voter intimidation claims.65


61 See 42 U.S.C. § 1985(3) (2012) (“If two or more persons in any State or Territory conspire or go in disguise on the highway or on the premises of another . . . .”).
62 See, e.g., Polidi v. Bannon, 226 F. Supp. 3d 615, 623 (E.D. Va. 2016) (“[A] plaintiff asserting a Section 1985 conspiracy must allege an agreement or a meeting of the minds by defendants to violate the claimant’s constitutional rights.”) (emphasis added) (internal quotation marks omitted).
64 52 U.S.C. § 10101(b) (2012).
65 See Voting Rights: Hearings on H.R. 6400 Before Subcomm. No. 5 of the House Comm. on the Judiciary, 89th Cong. 11 (1965) [hereinafter Hearings on H.R. 6400] (statement of Att’y Gen. Katzenbach) (“But perhaps the most serious inadequacy [of Section 131(b)] results from the practice of district courts to require the Government to carry the very onerous burden of
IV. SECTION 11(b) OF THE VOTING RIGHTS ACT & PLATFORM-BASED VOTER INTIMIDATION

In response to Section 131(b) of the 1957 Civil Rights Act, Attorney General Nicholas Katzenbach spearheaded enactment of the Voting Rights Act of 1965, which was designed to confront “the use of onerous, vague, unfair tests and devices enacted for the purpose of disenfranchising Negroes.”\(^{66}\) The Act put into place a sweeping statutory scheme, including a civil voter-intimidation provision that “represent[ed] a substantial improvement over [Section 131(b)].”\(^{67}\)

This provision, Section 11(b) of the Voting Rights Act of 1965, was designed to prohibit voter intimidation without forcing harmed individuals to demonstrate that violators, in fact, had the intent to discriminate.\(^{68}\) It reads:

No person, whether acting under color of law or otherwise, shall intimidate, threaten, or coerce, or attempt to intimidate, threaten, or coerce any person for voting or attempting to vote, or intimidate, threaten, or coerce, or attempt to intimidate, threaten, or coerce any person for urging or aiding any person to vote or attempt to vote . . . .\(^{69}\)

On its face, Section 11(b) avoids the flaws of earlier voter intimidation statutes by eliminating the triggering language of the KKK Act’s conspiracy requirement (“[i]f two or more persons . . . conspire”) and Section 131(b)’s mens rea requirement (“for the purpose of”). Yet relatively few litigants have proof of ‘purpose.’ Since many types of intimidation, particularly economic intimidation, involve subtle forms of pressure, this treatment of the purpose requirement has rendered the statute largely ineffective.”). Beyond Section 131(b)’s purpose requirement, a host of interpretative issues surfaced in lower court litigation of voter intimidation claims under the Civil Rights Act of 1957. Sherry Swirsky, Minority Voter Intimidation: The Problem that Won’t Go Away, 11 TEMP. POL. & CIV. RTS. L. REV. 359, 371 (2002) (“Courts applying [Section 131(b)] of the Act have reached conflicting conclusions on whether it reaches conduct by private individuals, which elections it covers, how much evidence of intimidation it requires, whether it may be enforced by private litigants, and if so, whether a private litigant must first exhaust state election board administrative remedies. Such inconsistency has posed an obstacle to meaningful enforcement of the provision.”).


\(^{67}\) Id. at 16.


\(^{69}\) 52 U.S.C. § 10307(b) (2012).
made use of Section 11(b)’s protections. Scholars have explained this dearth of caselaw as a function of how voter intimidation occurs in practice: Plaintiffs—individual and organizational—generally lack the incentive to litigate in-person incidents that only affect a few people on Election Day.

But platform-enabled voter intimidation presents a new kind of problem and a heightened scale of impact. Thus, the relative underdevelopment of Section 11(b)’s jurisprudence may present an opportunity to broaden voter protections. For Section 11(b) to reach and remedy modern, technologically advanced voter intimidation, three conditions must be met. First, platform-enabled voter intimidation must satisfy Section 11(b)’s statutory elements. Second, application of Section 11(b) to platform-enabled voter intimidation must not run afoul of the Constitution. Third, Section 11(b)’s relief adequately incentivize litigation against, and deter future violations by, perpetrators of voter intimidation.

A. Section 11(b)’s Statutory Reach to Platform-Enabled Voter Intimidation

Under a traditional legal analysis, evaluation of Section 11(b)’s applicability to platform-based voter intimidation comes before a broader inquiry into the First Amendment. This evaluation is highly context-dependent; a communication’s “intimidating,” “threatening,” or “coercive” character is informed by its surrounding technological-structural, socio-political, and social conditions. Because platforms can functionally obfuscate and distort these conditions, Section 11(b)’s statutory scheme can flounder at times. Still, its language seems suitable to reach at least some forms of platform-enabled voter intimidation.

To start, an open question exists as to whether Section 11(b)’s definition of “intimidation, threat, or coercion” reaches indirect and remote voter intimidation at all. Government attorneys Ben Cady and Tom Glazer argue convincingly that it does. They observe that methods of statutory interpretation, including ordinary usage, legislative history, and the

70 See Cady & Glazer, supra note 2, at 237–43.
71 Id. at 179–80.
72 Id. at 182.
73 See, e.g., Gomez v. United States, 490 U.S. 858, 864 (1989) (“It is our settled policy to avoid an interpretation of a federal statute that engenders constitutional issues if a reasonable alternative interpretation poses no constitutional question.”); see also Kevin Francis O’Neill, True Threats, FREE SPEECH CTR. AT MTSU: FIRST AMEND. ENCYCLOPEDIA, https://www.mtsu.edu/first-amendment/article/1025/true-threats [https://perma.cc/9DKH-W5LU] (“It is essential to distinguish between the court’s statutory analysis . . . and its constitutional analysis . . . .”).
74 See generally Cady & Glazer, supra note 2.
interpretive canon of in pari materia, all support a broad reading of Section 11(b)’s language that extends “not only [to] physical and economic coercion of voters, but also [to] a broader range of conduct that is intended to force prospective voters to vote against their preferences, or refrain from voting, through activity reasonably calculated to instill some form of fear.” This reading is a natural one that harmonizes with the traditionally liberal interpretation courts accord to remedial legislation.

However, even granting a broad interpretation of Section 11(b)’s text, a more difficult question arises concerning whether defendants must intend to instill fear in target audiences to be held liable under the Act. As discussed in Part IV supra, Section 11(b), by design, omits an express textual command requiring plaintiffs to demonstrate a defendant’s purpose in proving voter-intimidation claims. Attorney General Katzenbach took great care to exclude the phrase “for the purpose of” from Section 11(b) to sidestep the shortfalls of Section 131(b). This decision was specifically calculated to relieve litigants of the “very onerous burden of proof” of showing a defendant’s motives. As Katzenbach observed, “defendants would be deemed to intend the natural consequences of their acts.”

Nonetheless, Katzenbach’s design may have failed to foreclose a mens rea requirement from yet again creeping into a federal civil voter-intimidation statute. Legislative history carries negligible weight in today’s textualist courts, especially when plain language is silent or contradictory. And Section 11(b) does not disclaim the need for victims to demonstrate a violator’s “purpose,” nor do the words “intimidate,” “threaten,” and “coerce” imply that courts should look only to the natural consequences of a defendant’s conduct. On the contrary, Black’s Law Dictionary defines “intimidate” as “unlawful coercion, extortion, duress, [or] putting in fear,” and notes that “[s]uch fear must arise from the willful conduct of the accused.” It also recognizes that a

75 Id. at 201 (internal quotation marks omitted) (quoting U.S. ELECTION ASSISTANCE COMM’N, ELECTION CRIMES: AN INITIAL REVIEW AND RECOMMENDATIONS FOR FUTURE STUDY 14 (Dec. 2006)).
77 See supra Part IV (“Section 11(b) avoids the flaws of earlier voter intimidation statutes by eliminating the triggering language of . . . Section 131(b)’s intentionality requirement.”).
79 Id.; see also H.R. REP. NO. 89-439, at 30 (1965) (“The prohibited acts of intimidation need not be racially motivated; indeed, unlike [Section 131(b)] (which requires proof of a ‘purpose’ to interfere with the right to vote) no subjective purpose or intent need be shown.”).
81 Intimidate, BLACK’S LAW DICTIONARY (6th ed. 1979) (emphasis added).
person is guilty of coercion “[only] if [they act] with purpose to unlawfully restrict another’s freedom of action to his detriment . . . .”

In turn, counter to Katzenbach’s design, some courts have required plaintiffs to demonstrate a defendant’s intent to “intimidate, threaten, or coerce” voters under Section 11(b). This reading presents obstacles to plaintiffs pressing claims against platform-enabled speakers premised on a theory of strict liability. Ordinarily, in the absence of a defendant’s stipulation of intent, courts would look to readily discernable facts that support a finding of a defendant’s mens rea. But platforms can obfuscate these indicators. Innocently worded messages can be designed to impose crippling intimidation on target audiences when speakers anticipate that their followers will spread their words far and wide. And other platform speakers can join, color, and distort an original speaker’s message through participatory application programming interfaces and “resharing.” In practice, the fog of a platform’s design and algorithmic logic can mask underlying motives.

Fortunately, there are methods for cutting through this haze—for imputing “purpose” to a platform-enabled speaker. The first is to borrow the “reasonable speaker” test from cases that deal with federal threat statutes and the “true threat” exception to free speech protections. Because these cases deal with conduct similar to and, at times, overlapping with the conduct at issue under Section 11(b), they provide a reasonable baseline for ascertaining a speaker’s intent in the absence of an express admission or some other form of clear evidence.

82 Coercion, BLACK’S LAW DICTIONARY (6th ed. 1979) (emphasis added).
83 See, e.g., Olagues v. Russoniello, 770 F.2d 791, 804 (9th Cir. 1985) (citing United States v. McLeod, 385 F.2d 734, 740–41 (5th Cir. 1967)) (“[T]he organizations’ claims under the Voting Rights Act against these officials do not appear to have merit. If the search of voting records intimidated bilingual voters, such intimidation would satisfy only one part of a two-pronged test for violations of [Section 11(b)]: the voters and organizations were intimidated, but the officials did not intend to intimidate.”). Not all courts have required intentionality under Section 11(b). See League of United Latin Am. Citizens - Richmond Region Council 4614 v. Pub. Int. Legal Found., No. 1:18-CV-00423, 2018 WL 3848404, at *4 (E.D. Va. Aug. 13, 2018) (“Defendants’ reference to nonbinding case law that reads specific intent and racial animus requirements into § 11(b) is also unpersuasive. These cases trace back to United States v. McLeod [385 F.2d 734, 738 (5th Cir. 1967)], which, in fact, adjudicated claims brought under the 1957 Civil Rights Act. Therefore, in the absence of plain statutory text, statutory history, or binding case law to the contrary, the Court does not find that a showing of specific intent or racial animus is required under § 11(b).”).
84 Cf. United States v. Bruce, 353 F.2d 474 (5th Cir. 1965) (observing that the intent element of Section 131(b) was satisfied when a defendant invoked his right to exclude from his property a black insurance collector, who had previously been given free access to the property, when the defendant discovered that the plaintiff was undertaking efforts to register black voters).
85 CITRON, supra note 6, at 66–67.
In the context of other statutes, courts that apply the “reasonable speaker” test impute intent to speakers who make statements “a reasonable person would foresee . . . would be interpreted by those to whom the maker communicates [them] as a serious expression of an intention to inflict bodily harm.”\textsuperscript{86} A prominent deployment of the “reasonable speaker” test occurred in the Ninth Circuit’s en banc ruling in \textit{Planned Parenthood of the Columbia/Willamette, Inc. v. American Coalition of Life Activists.}\textsuperscript{87} There, four abortion providers brought suit under the federal Freedom of Access to Clinic Entrances Act, “claiming that they were targeted with threats by several groups of anti-abortion activists.”\textsuperscript{88} The activists had printed “Deadly Dozen” posters, which identified and listed the names, addresses, and photographs of the plaintiff-doctors.\textsuperscript{89} Some of the posters also read “GUilty” and “WANTED.”\textsuperscript{90} After a Ninth Circuit panel found that the posters were protected speech and could not lose their protections by the context in which they were displayed, the circuit granted rehearing en banc and reversed.\textsuperscript{91} The court defined “true threat” as “a statement which, in the entire context and under all the circumstances, a reasonable person would foresee would be interpreted by those to whom the statement is communicated as a serious expression of intent to inflict bodily harm upon that person.”\textsuperscript{92}

Through this holding, the court expressly disavowed the view that defendants must “intend to, or be able to carry out . . . threat[s]” to be held liable for their statements.\textsuperscript{93} Rather, it observed that “the only intent requirement for a true threat is that the defendant intentionally or knowingly communicate the threat.”\textsuperscript{94} In the end, the anti-abortion activists were found liable because reasonable speakers in their position would foresee that the doctors identified by their posters would interpret their naming as a serious expression of the speakers’ intent to inflict bodily harm upon them.

Adapted to platform-enabled voter intimidation under Section 11(b), a modified version of the “reasonable speaker” test would ask whether reasonable users of a platform would foresee that their message would have an intimidating, threatening, or coercive effect on the voting behavior of its recipients. Under this approach, two factors would guide the inquiry into a

\textsuperscript{86} \textit{See, e.g.}, United States v. Kosma, 951 F.2d 549, 557 (3d Cir. 1991).
\textsuperscript{87} \textit{Planned Parenthood of the Columbia/Willamette, Inc. v. Am. Coal. of Life Activists}, 290 F.3d 1058 (9th Cir. 2002) (en banc); \textit{see also} Paul T. Crane, “True Threats” and the Issue of Intent, 92 VA. L. REV. 1225, 1248–52 (2006).
\textsuperscript{88} \textit{Am. Coal. of Life Activists}, 290 F.3d at 1062.
\textsuperscript{89} \textit{Id.} at 1064–65.
\textsuperscript{90} \textit{Id.} at 1064.
\textsuperscript{91} \textit{See generally id.}
\textsuperscript{92} \textit{Id.} at 1077.
\textsuperscript{93} \textit{Id.} at 1075.
\textsuperscript{94} \textit{Am. Coal. of Life Activists}, 290 F.3d at 1075.
speaker’s intent: (1) the content of the speaker’s message and (2) the background conditions surrounding the speech. This approach is imperfect, but it captures some platform-enabled voter intimidation without sweeping in an excess of benign online user activity not calculated by speakers to impinge upon others’ voting rights.

An application of the test to several of the examples of intimidation identified in Part I is illustrative. Mac McDonald’s 2016 Election Day Facebook post reads, “I Pray Trump wins as we veterans and militia are ready to take our country back. You civilians either follow us or get out of the way. Your only other option is to be a qualified target.”  

The first factor—the content of McDonald’s message—is no doubt threatening. It can be read as offering an ultimatum: Vote for Trump or risk bodily insecurity. Without more, a reasonable speaker would foresee that the statement would have an intimidating effect on its recipients. Compared to a subjective standard of intent, it would not matter if McDonald’s friends and family testified that he often used the term “qualified target” to mean “someone who is ignorant and needs education.” The message speaks for itself.

On the other hand, the second factor—the message’s background conditions—could just as easily undermine the intimidation charge under the modified “reasonable speaker” test. Consider the social and technological-structural characteristics of the platform on which the post is made. McDonald directs his message to his “connections” on Facebook. If McDonald’s Facebook network is entirely comprised of ideologically and politically aligned individuals with similar backgrounds, and if McDonald’s profile-viewing permissions are set to private, meaning that only his Facebook connections can see his posts, then a reasonable speaker in McDonald’s position would foresee that his macho bravado would prompt grunted laughter among his audience, rather than fear. These conditions would operate to offset the message’s threatening content under the modified “reasonable speaker” test and would most likely render any resulting voter intimidation unintentional on McDonald’s part under Section 11(b).

For a second example, consider Dillian Billiot’s 2016 Election Day Facebook post. Beside a photograph of an AR-15 rifle and a pistol together on a mattress, the post reads “Where’s all my Militia brothers? I’m on standby ready to defend our constitution and fight for our freedom. I’ll volunteer for this country any day.”  

Here, the first factor—the message’s content—sends mixed signals: The text is innocuous, but the image forebodes violence against voters. The second factor—the message’s background conditions—is also potentially ambiguous. Assume that Billiot is an Internet celebrity with a large

---

95 Appendix 1.2.
96 Appendix 1.1.
following and that his profile-viewing permissions are set to public, such that Facebook’s algorithm predictably spreads his messages to a wide audience. Perpetrators of mass shootings are seldom public figures. A reasonable speaker in Billiot’s position, were Billiot a celebrity, might therefore have reason to believe that readers of his message would be unlikely to interpret it as a serious threat of violence. On the other hand, if Billiot’s message were spread far and wide across the web, there is a greater chance that it would be clicked and shared by his connections for the purpose of intimidating their audiences—audiences that lack familiarity with Billiot’s “celebrity” status. The likelihood of this “derivative intimidation” increases with the scale of Billiot’s network. As a result, a reasonable speaker in his position could foresee that his message might be “weaponized” by one of his connections as a tool of platform-enabled voter intimidation. In short, the background conditions of Billiot’s speech can conflict under the modified “reasonable speaker” test, complicating its ease of administration.

But courts are no strangers to complex balancing tests, which are commonplace throughout many areas of the law.\footnote{See, e.g., Manson v. Brathwaite, 432 U.S. 98, 115–17 (1977) (employing a multi-factor balancing test to determine the reliability of out-of-court identifications); Gordon v. United States, 383 F.2d 936, 939–41 (D.C. Cir. 1967) (using numerous factors to balance probative value and prejudicial effect of prior convictions offered for witness impeachment).} Even with anonymous posts on 4chan, a court or a jury could ascertain whether a reasonable speaker in the position of the poster should have foreseen that the speaker’s message would have an intimidating, threatening, or coercive effect on the voting behavior of the message’s recipients. These fact-finders would simply look to the language used in the post, the comments to which the post responds, the responses the post elicits, and the discussion board to which the post is submitted as evidence.

Even beyond its suitability for determining the intent underlying platform-enabled speech, the modified “reasonable speaker” test offers the benefit of demonstrating that a victim of platform-enabled voter intimidation has reasonably been impacted by the speech in question. Of the few Section 11(b) claims that have been brought to date, several have been dismissed because plaintiffs were unable to show that the alleged intimidation had a reasonable, tangible impact on actual voters.\footnote{United States v. Edwards, 333 F.2d 575, 579 (5th Cir. 1964); Brooks v. Nacrelli, 331 F. Supp. 1350, 1353 (E.D. Pa. 1971); Cady & Glazer, supra note 2, at 195.} A finding of intent under the modified “reasonable speaker” test would fulfill this element. It would not matter if alleged victims were temperamentally timid or overly sensitive. A finding of a speaker’s intent would suffice.

Looking beyond the modified “reasonable speaker” test, there are several other options for imputing intent to a platform-enabled speaker for
purposes of Section 11(b) liability, though none that are equally effective. For one, courts could import the “reasonable listener” test from the same line of federal threat-statute cases. Used in a minority of circuits, this test asks whether “an ordinary reasonable recipient who is familiar with the context of [a statement] would interpret it as a threat of injury.”99 Adapted to platform-enabled voter intimidation under Section 11(b), a modified version of this test would ask whether an ordinary recipient of an online communication, familiar with its context, would be intimidated, threatened, or coerced by exposure to it. And like the modified “reasonable speaker” test, it would similarly factor (1) the content of the speaker’s message, and (2) the background technological-structural, socio-political, and social conditions surrounding platform-enabled speech.

In terms of its unique benefits, the modified “reasonable listener” test has the advantage of expanding intent in narrow circumstances by measuring it at the time a recipient encounters an intimidating post, rather than at the time of its initial making. Rehashing the example of McDonald’s Facebook post, assume that McDonald rightly calculates that his message will not intimidate, threaten, or coerce any of his connections in his closed, private network. If, unexpectedly, one of McDonald’s connections reshares his post to intimidate a target outside of McDonald’s network, McDonald could still be liable under the modified “reasonable listener” test, provided that the target outside his network is familiar with the context of his statement and is reasonably intimidated by it.

But this raises the question: How much familiarity must an ordinary recipient of a platform-enabled communication have with the communication’s context for intent to attach to the speaker under the “reasonable listener” test? When the Fourth Circuit first adopted the standard in 1973, it determined that a personal history between the maker and the recipient of a particular threatening communication was a sufficient nexus to establish “context familiarity.”100 But platform users frequently do not share a “personal nexus” with other users that view their posts. Platform-enabled voter intimidation, typically anonymous and generalized in character, would therefore regularly skirt Section 11(b) liability.

Perhaps in recognition of this limitation, courts in recent cases have embraced a more lenient standard of “context familiarity.” In United States v. Turner, the Second Circuit found that three Seventh Circuit judges were

99 United States v. Maisonet, 484 F.2d 1356, 1358 (4th Cir. 1973) (emphasis added); see also Crane, supra note 87, at 1246 (describing the “reasonable speaker test” as the majority rule and noting that the “reasonable listener test” has only been adopted in four circuits: the Second, Seventh, Eighth, and Eleventh).
100 See Maisonet, 484 F.2d at 1358 (“Maisonet had been sentenced to prison by the judge to whom he addressed the letter.”).
sufficiently familiar with the context of a blog post that threatened them to render the poster liable for intimidation, even though “[n]one of the three judges had ever heard of [the author] before reading [the] post.”\footnote{101} The court found that the “seriousness of the threat” and the author’s “[o]ther posts readily accessible on the blog . . . provided context from which a reader might infer [the author’s] intentions in writing the post.”\footnote{102} Under this application of the standard, an ordinary speaker encountering an anonymous threat on a platform could reasonably take stock of the seriousness of the threat (the content of the speech) and other readily accessible posts and information volunteered by the speaker (the background conditions surrounding the speech). Taking the example of @NeilTurner’s tweet detailed in Part I, a court would look to whether an ordinary voter who encountered the professionally doctored photo on Twitter and had access to @NeilTurner’s previous tweets would feel intimidated, threatened, or coerced by exposure to the bot’s post. If an ordinary reader would find the photograph credible and be deterred from voting out of fear of being targeted by immigration officials, the architect behind @NeilTurner could be held liable in a Section 11(b) “intent” jurisdiction.

But there are many other examples in which the modified “reasonable listener” test vastly underperforms the modified “reasonable speaker” test in terms of its reach to wrongful conduct. Assume, for example, that Mac McDonald \textit{intends} to intimidate voters in coordination with other users. Provided that he sets his profile-viewing permissions to “network-viewable,” hundreds of his connections could reshare his “qualified target” remark to thousands of Facebook users, none of whom would have access to McDonald’s own posting history or profile information. Given McDonald’s ability to obscure the background conditions surrounding his speech, targets of “derivative intimidation” would not be able to achieve “context familiarity” with his original communication under the modified “reasonable listener” test. Therefore, such targets would be unable to hold McDonald liable in an “intent” jurisdiction under Section 11(b).

Although plaintiffs could try to hold users that reshare McDonald’s post liable, it is also not at all clear that the content of the original post could be attributed to these “derivative defendants.” There are plenty of reasons to reshare a post: to ratify it, to poke fun at it, or to draw critical attention to it. It bears repeating that a platform user’s intent is not always facially evident, even with respect to a retweet or a repost. Aside from this issue, ephemeral bots may lack platform histories entirely, making it difficult for plaintiffs to achieve any level of context familiarity at all. All told, the modified “reasonable

\footnote{101} See United States v. Turner, 720 F.3d 411, 416 (2d Cir. 2013).
\footnote{102} \textit{Id.} at 416, 422.
“reasonable listener” test would allow many instances of platform-enabled voter intimidation to occur with impunity, given the unique communicative properties of social media sites and other third-party online forums.

Another standard for imputing an online commenter’s intent, more extreme and overinclusive than employing the modified “reasonable listener” test, would be to import tort liability frameworks to the intent inquiry under Section 11(b). Strict liability, often rationalized on economic grounds, could be used as a libertarian instrument.103 Where platform-enabled actors cause harm to voters through use of compulsion, force, or the creation of a “dangerous condition,” rights-based norms could justify imputing intent automatically.104 In theory, this approach would ensure that speakers on Internet platforms who play some role in platform-enabled voter intimidation—whether by posting comments, spreading them, or “liking” them—could be enjoined through use of Section 11(b) in “intent” jurisdictions. Similarly, joint and several liability, which allows plaintiffs to hold independent tortfeasors liable for a single, theoretically divisible but practically indivisible harm, would ensure that plaintiffs would be able to bring suit against any defendant who played a role in the dissemination of an intimidating communication105 But both of these tort-liability models are likely to expand the scope of intent far beyond the tolerable limits of the First Amendment.106 A quick look at potential consequences of this move reveals its risks of overinclusiveness. Actors like the Southern Poverty Law Center, attempting to draw negative attention to reprehensible voter intimidation by reshar ing it, could be held liable. Similarly, the complicated algorithmic designs of platforms could render liable those third-party users that merely view an intimidating post, as “views” often contribute to a post’s visibility and virality. Such an expansive assignment of intent for purposes of Section 11(b) would disproportionately extend liability without fault.

In sum, shortfalls of the modified “reasonable listener” test and the tort-liability approach reveal the comparative strengths of using the modified “reasonable speaker” test to assign intent to platform-enabled users for purposes of Section 11(b) liability in intent jurisdictions. The modified “reasonable listener” test permits bad actors to skirt liability by narrowly sharing intimidating content with a wink and a nod. The tort-liability approach sweeps in bad actors, but also reaches vast amounts of innocuous expressive activity. The modified “reasonable speaker” test suffers from neither of these defects.

104 Id. at 205.
105 See id. at 139.
106 See infra Part IV.B.
B. Overcoming Constitutional Limitations

Even assuming a Section 11(b) suit against a perpetrator of platform-enabled voter intimidation could satisfy all the requisite statutory elements, the Constitution imposes its own limitations on private litigants’ use of federal civil voter-intimidation statutes to proscribe expression. For one, restrictions of platform-enabled speech would constitute content-based regulations under the First Amendment, and thus be subject to strict scrutiny in the absence of some categorical exception.\(^{107}\) To survive strict scrutiny, use of Section 11(b) must be “narrowly tailored to serve a compelling state interest.”\(^{108}\) The Supreme Court has recognized “a compelling interest in protecting voters from confusion and undue influence.”\(^{109}\) The key inquiry, then, is whether litigants’ use of Section 11(b) to proscribe platform-enabled voter intimidation is narrowly tailored to this compelling interest.

Generally, courts find speech regulations narrowly tailored and, thus, constitutional when they (1) advance the state interest, (2) are the least restrictive means available to advance the interest, (3) are not overinclusive, and (4) are not underinclusive.\(^{110}\) Under these parameters, use of Section 11(b) should generally pass constitutional muster. Taking Mac McDonald’s “qualified target” remark as an example, the post’s removal via injunction would need to be necessary to avoid undue influence over voters, removal would have to be the least restrictive means to avoid undue influence, removal would have to avoid restricting McDonald’s other protected speech, and removal would have to address all of McDonald’s speech that exerts undue influence over voters. All these requirements can arguably be met.

That’s not to say that all platform-enabled voter intimidation can be proscribed without running afoul of the First Amendment. In the case of @NeilTurner, even if the immigrant-voter tweet could have been lawfully removed when it was first posted, the First Amendment calculus changed when the photo was picked up and redistributed by other users and major news


\(^{109}\) Burson, 504 U.S. at 199.

outlets. At that point, removal could not diminish the post’s effect on voter behavior. An injunction would thus fail narrow tailoring.\(^{111}\)

Where applications of Section 11(b) do not survive the Court’s narrow-tailoring requirement, it is also unlikely that the First Amendment exception for “true threats” can salvage their constitutionality. In the original “true threat” case, *Watts v. United States*, the Supreme Court found that a 1917 federal law that prohibited “any person from knowingly and willfully making any threat to take the life of or to inflict bodily harm upon the President of the United States” could not be used to prosecute a defendant who, reacting to his Vietnam War draft order, announced:

They always holler at us to get an education. And now I have already received my draft classification as 1-A and I have got to report for my physical this Monday coming. I am not going. If they ever make me carry a rifle the first man I want to get in my sights is L.B.J.\(^{112}\)

In determining that the defendant’s speech was protected and not a “true threat,” the Court in *Watts* focused on the surrounding facts: that his declaration was made during a political debate, that his threat was conditional, and that the audience reacted to his statements with laughter.\(^{113}\) Given this context, the Court found that Watts’s speech—which mirrors much of the platform-enabled voter intimidation identified in Part I—aligned with “the background of a profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open . . .”\(^{114}\) This principle all but commands the constitutional protection of most platform-enabled voter intimidation—the government’s removal of which would not be narrowly tailored to the compelling state interest of protecting voters from confusion and undue influence.

Separate from the First Amendment limitations of narrow tailoring and the true threat exception, there also remains a question concerning Congress’s

\(^{111}\) Furthermore, neither the post’s falsity nor its intimidating character could suffice to bring it outside of the First Amendment’s protections. *See* United States v. Alvarez, 567 U.S. 709, 720 (2012) (explaining a certain statute’s “prohibition on false statements made to Government officials, in communications concerning official matters, does not lead to the broader proposition that false statements are unprotected when made to any person, at any time, in any context”); Virginia v. Black, 538 U.S. 343, 360 (2003) (“Intimidation in the constitutionally proscribable sense of the word is a type of true threat, where a speaker directs a threat to a person or group of persons *with the intent of placing the victim in fear of bodily harm or death.*”) (emphasis added).


\(^{113}\) *Id.* at 707.

\(^{114}\) *Id.* at 708.
power to regulate purely private voter intimidation. The Elections Clause provides that “the times, places and manner of holding elections for Senators and Representatives, shall be prescribed in each state by the legislature thereof; but the Congress may at any time by law make or alter such regulations . . .”115 The Supreme Court has interpreted this language to vest Congress with broad authority to regulate federal elections, including those in which state and local candidates run.116 But if the Elections Clause is the source of power for the Voting Rights Act, it is doubtful that Section 11(b) can reach private speech connected to purely state and local political contests.117 The Act’s own language seems to recognize this limitation. It only extends protections to voters for:

... any candidate for the office of President, Vice President, presidential elector, Member of the United States Senate, Member of the United States House of Representatives, Delegate from the District of Columbia, Guam, or the Virgin Islands, or Resident Commissioner of the Commonwealth of Puerto Rico.118

However, even though this constitutional limitation theoretically restricts the reach of Section 11(b), it seems of limited practical consequence in a world in which state and local elections are almost always conducted concurrently with

116 See Arizona v. Inter-Tribal Council of Arizona, 570 U.S. 1, 8–9 (2013); Ex parte Yarbrough, 110 U.S. 651, 661–62 (1884).
117 There are exceptions to this rule for “major parties and other people or groups who are either performing traditional public functions or entangled with the state.” See Tokaji, supra note 107, at 103–04 (citing Reitman v. Mulkey, 387 U.S. 369 (1967) and Smith v. Allwright, 321 U.S. 649 (1944)).
118 52 U.S.C. § 10307(c) (2012). Some scholars have suggested that the Necessary and Proper Clause could help Congress fill this “coverage gap” in Section 11(b). See Cady & Glazer, supra note 2, at 212 n.254 (quoting Foster v. Love, 522 U.S. 67, 71 n.2 (1997) for the proposition that the Elections Clause “gives Congress comprehensive authority to regulate the details of elections including the power to impose the numerous requirements as to procedure and safeguards which experience shows are necessary in order to enforce the fundamental right involved.”) (emphasis added) (internal citations omitted). But the Supreme Court has only ever held that the Necessary and Proper Clause “safeguard[s] the right of choice by the people of representatives in Congress secured by § 2 of Article I.” See United States v. Classic, 313 U.S. 299, 319–21 (1941) (emphasis added); see also Tokaji, supra note 107, at 104. But see Shelby Cty. v. Lynch, 799 F.3d 1173, 1181 (D.C. Cir. 2015) (Tatel, J., concurring) (“That Congress may enforce the Amendments only by ‘appropriate’ legislation, the County insists, means that the enforcement provisions guarantee ‘the constitutional right of sovereign States . . . to regulate state and local elections as they see fit.’ But this claim finds no support in the constitutional text.”) (citations omitted).
federal elections for the ease of administration.\textsuperscript{119} Even if states did bifurcate these elections, all the examples of platform-enabled voter intimidation identified in this Note pertain to national-level elections.\textsuperscript{120} That makes sense, because the Internet’s global reach makes it particularly suitable for discussing issues of broad national and global import. Constitutional limits on federal regulatory power over private voter intimidation are therefore unlikely, in practice, to bar private litigants’ use of Section 11(b) to deter purely private platform-enabled voter intimidation.

C. Obtaining Adequate Relief Under Section 11(b)

Assuming victims of and objectors to platform-based voter intimidation can use Section 11(b) without overstepping statutory and constitutional limitations, does the relief available under Section 11(b) sufficiently incentivize litigation and deter future violations? Other civil voter-intimidation statutes authorize compensatory damages for “part[ies] so injured or deprived.”\textsuperscript{121} But similar language is absent from Section 11(b).\textsuperscript{122} Courts have thus drawn the line at declaratory and injunctive relief, removing a key incentive for bringing private suit.\textsuperscript{123}

Fortunately, Section 11(b)’s provisions have not been reduced to a total nullity. In 1976, Congress amended the Voting Rights Act with 52 U.S.C. § 10310(e), which provides that, “In any action or proceeding to enforce the voting guarantees of the fourteenth or fifteenth amendment, the court, in its discretion, may allow the prevailing party . . . a reasonable attorney’s fee, reasonable expert fees, and other reasonable litigation expenses as part of the

\textsuperscript{119} See Arizona v. Inter-Tribal Council of Arizona, 570 U.S. 1, 41 (2013) (Alito, J., dissenting) (“[T]he Elections Clause’s default rule helps to protect the States’ authority to regulate state and local elections. As a practical matter, it would be very burdensome for a State to maintain separate federal and state registration processes with separate federal and state voter rolls. For that reason, any federal regulation in this area is likely to displace not only state control of federal elections but also state control of state and local elections.”).

\textsuperscript{120} Appendix 1.1–3.3.

\textsuperscript{121} 42 U.S.C. § 1985(3) (2012).

\textsuperscript{122} See 52 U.S.C. § 10307(b) (2012).

\textsuperscript{123} See Allen v. State Bd. of Elections, 393 U.S. 544, 555 (1969); Olagues v. Russoniello, 770 F.2d 791, 805 (9th Cir. 1985). In Olagues, the Ninth Circuit observed that “[t]he legislative history [of the Voting Rights Act] nowhere suggests any action for damages, but instead observes that a private litigant is entitled to ‘the same remedy’ as the Attorney General . . . that [legislative] history points out that ‘[t]he sole consequence’ of the provision for a private cause of action under the Act is to broaden the scope of equitable relief which may be requested’ . . . .” 770 F.2d at 805 (citations omitted) (quoting S. REP. NO. 295, 94th Cong., 1st Sess. 39–43, \textit{reprinted in} 1975 \textit{U.S. CODE CONG. & AD. NEWS} 774, 806–10).
costs.”124 This language provides at least some financial impetus for nonprofit litigators to secure compliance with the law’s strictures.125

But attorney’s fees, without more, hardly seem adequate to deter the rising tide of diffuse voter intimidation spread across a vast array of mostly anonymous platforms. Few Section 11(b) actions have been brought to date, as individual voters lack the financial incentives to sue.126 Meanwhile, because voter intimidation has historically been isolated and difficult to prove, advocacy organizations have preferred to challenge conduct and laws with more tangible, far-reaching effects on constituencies.127 However, the proliferation of voter intimidation on Internet platforms may somewhat change this calculus. Organizational plaintiffs can reasonably envision a greater return on investment by containing the rippling effects of “information cascades”—especially where deep-pocketed organizational speakers like PILF shed the cloak of anonymity to spread their message “far and wide.”128

Still, it seems unlikely that any plaintiff—individual or institutional—would be willing to bear the burden of locating and extinguishing voter intimidation at the hands of individuals that surfaces on platforms.129 Even if this voter intimidation was easily ascribable to a specific speaker, most of it would occur on Election Day, making it virtually impossible to enjoin in a timely manner.130 Any judicially-imposed relief would therefore be substantively moot, and could not be litigated as an issue “capable of repetition, yet evading review” given the unique, factually-specific character of the intimidation.131

126 Cady & Glazer, supra note 2, at 179; see also Delegates to Republican Nat’l Convention v. Republican Nat’l Comm., No. SACV 12-00927 DOC, 2012 WL 3239903, at *11 (C.D. Cal. Aug. 7, 2012) (“The Court has found only four dozen cases discussing [Section 11](b) of the Voting Rights Act.”).
128 See CITRON, supra note 6, at 66–67; see generally LULAC Complaint, supra note 31.
129 See, e.g., Appendix 1.1, 1.3–1.6, 3.1–3.2.
130 See CITRON, supra note 6, at 221 (“Perpetrators cannot be sued or indicted if they cannot be identified.”); see also Appendix 1.1–3.3.
131 Compare the highly variable and fact-specific occurrence of voter intimidation with the consistent and relatively uniform issue of pregnancy termination addressed in Roe v. Wade, 410 U.S. 113, 125 (1973).
V. SOLUTIONS BEYOND EXISTING LAW

As Part IV, supra, explains, current law is ill-equipped to address most forms of platform-enabled voter intimidation. Unfortunately, prospects for alternative low-risk solutions are equally bleak. Traditional agency structures and hybrid administrative paradigms are poorly matched to the dynamic, fast-paced, and quasi-expressive nature of the problem. Platform self-policing offers some promise but may not be reliable or trustworthy given profitability considerations. And fundamental alterations to the existing electoral process would be expensive and might not blunt the effects of platform-enabled voter intimidation.

What would change look like under our existing election system? First, Congress could charge an agency with the mandate of stymying speech that “intimidates, threatens, or coerces” voters on social media sites and third-party forums. This agency’s tasks could include setting binding “voter intimidation standards” via rulemaking, providing content-moderation guidance to platform administrators via policy statements, and issuing interpretative decisions in response to direct petitioning from individual platforms. Lawmakers could also furnish the agency with online-content “take-down” authority, along with the ability to initiate enforcement actions against regulated entities that deliberately shirk compliance. In theory, this approach would afford significant advantages over the status quo—namely accountability, the enshrinement of national legal values, and nonarbitrary enforcement.

But from where would the political will to create such an agency derive? The Trump administration has shown more interest in stripping the power of independent agencies than creating new ones. Furthermore, there is no existing “heir apparent” administrative body to which the directive could be tasked. The Federal Election Commission’s mandate has always been limited to campaign finance, and it already has fewer enforcement resources than problems to address.

---

132 See infra notes 135–46 and accompanying text.
133 See infra notes 147–58 and accompanying text.
134 See infra notes 161–72 and accompanying text.
Even if the predicate political will did exist, it is unlikely that a traditional command-and-control regulatory model, a vestige of the New Deal era, could realistically pace dynamic, fast-moving platform-enabled speech. As Orly Lobel observes in her influential scholarship on regulatory limitations, traditional “command-and-control” processes are, by their nature, static and ossified, and generally employ “reactive, defensive, ex post” procedural responses to undesired conduct. For platform-enabled voter intimidation that sprouts rapidly and unpredictably throughout the web on Election Day, slow-moving, after-the-fact government regulation hardly seems like an adequate remedial instrument.

Still, there has been a substantial improvement over the past few years in the technological capacity of the private sector to identify and track certain bot-enabled voter-intimidation activities. The use of data-mining techniques and machine-learning tools has greatly eased this detection burden over the last few years. Setting aside budgetary constraints, this technology could easily be ported over from the private contractor base to government agencies. However, not all voter intimidation is perpetrated by malicious
actors who emit identifiable network signatures. And not all automated tools are perfectly accurate. In practice, only so much platform-enabled voter intimidation can be flagged by algorithmic software suites on government terminals.

The practical limitations of automated intimidation-tracking software casts doubt on the constitutional validity of charging a traditional regulatory agency with thwarting platform-enabled speech. Even if Congress, in its enabling legislation, cabined the agency’s “take-down” and enforcement discretion with procedural checks to minimize the censorship of protected speech on platforms, the agency could not feasibly eradicate all platform-enabled voter intimidation. Such a failure would render its legal sanction underinclusive. It would also inevitably restrict some protected “core political speech” incident to its directive—an “overinclusive” regulation anathema to what Justice Clarence Thomas has called “the primary object of First Amendment protection.” These inadequacies would presumably render such an anti-intimidation protocol invalid under the Court’s existing “narrow tailoring” prong for First Amendment strict scrutiny.

Given these shortfalls, one might alternatively argue for the implementation of a hybrid institutional structure that uses what Professor Henry H. Perritt Jr. describes as “broad public law frameworks within which private regulatory regimes work out the details.” This “new governance” model could address many of the limitations of the traditional regulatory approach by transferring certain standard-setting and enforcement responsibilities to entities that can most effectively harness “participation, collaboration, active citizenship, proliferated production, dynamic learning, and adaptability.” Under this approach, a hybrid entity could be comprised of strategic and technical stakeholders from social media sites and third-party

---

144 See Volokh, supra note 110, at 2420.
146 See supra note 110 and accompanying text. Concededly, the Court, on occasion, has afforded legislatures considerable leeway when analyzing whether statutes that guard against corruption of the electoral process are narrowly tailored. See Burson v. Freeman, 504 U.S. 191, 208–09 (1992) (“[T]his Court never has held a State to the burden of demonstrating empirically the objective effects on political stability that are produced by the voting regulation in question. Elections vary from year to year, and place to place. It is therefore difficult to make specific findings about the effects of a voting regulation. Moreover, the remedy for a tainted election is an imperfect one. Renunning an election would have a negative impact on voter turnout.”) (citations omitted).
147 Perritt, supra note 135, at 250.
forums subject to its jurisdiction. It could exercise direct content take-down authority, or, alternatively, wield indirect notification tools to subject sites hosting noncompliant speech to legal sanction. A governmental appellate body, judicial or otherwise, could offer review of its initial determinations.\(^{149}\) And the hybrid entity could work in tandem with a governmental authority to define standards for government-instituted enforcement actions against platforms that repeatedly aid or abet voter intimidation.\(^{150}\)

However, like the traditional regulatory agency model, this approach would ultimately suffer from constitutional infirmities—both under the First Amendment and under the non-delegation doctrine, which proscribes delegation of legislative power to private entities.\(^{151}\) As a result, this hybrid institution would most likely need to be formed through voluntary commitments from leading industry players. Such an arrangement is not without precedent. In 2016, Facebook, Microsoft, YouTube, and Twitter entered into a voluntary agreement with the European Commission to remove the most “extremist” speech on their respective platforms within twenty-four hours of its posting.\(^{152}\) As part of this agreement, the companies developed a collective database for flagging and removing extremist content.\(^{153}\) But even hybrid institutions operating with implicit state sanction present inherent risks. Professor Jodi Short notes that implementations that place private actors in nominally traditional administrative roles can provide cover for devolution of the government’s central public-interest function.\(^{154}\) As a result, industry capture, factionalism, and interest-group pressure can cross over into

\(^{149}\) Perritt, *supra* note 135, at 303.

\(^{150}\) The outcomes of these enforcement actions could mirror the penalties available under Section 5 of the FTC Act. See 15 U.S.C. § 45 (2012).


\(^{154}\) See generally Jodi Short, *The Paranoid Style in Regulatory Reform*, 63 HASTINGS L. J. 633 (2012); see also ROGER G. NOLL, REFORMING REGULATION 40–43, 46 (1971) (expressing skepticism of the role for private actors in traditional regulatory contexts).
peripheral regulatory contexts.\textsuperscript{155} The stakes of these risks are elevated by the extensive intellectual capital already invested by the administrative agencies “in legitimizing the constitutionally suspect ‘headless fourth branch’ of government.”\textsuperscript{156}

How, then, can such an entity be formed without leaving the fox in charge of the henhouse? Jody Freeman, Director of the Environmental and Energy Law Program at Harvard Law, has asserted that legally enforceable contracts, the presence of powerful independent professionals within private organizations, the background threat of regulation by an agency, professional norms, and informal sanctions all work to deter the corrupting influence of private interest.\textsuperscript{157} Nevertheless, these proposed solutions seem unlikely to provide accountability in the context of government oversight of private regulation of platform-enabled voter intimidation. The problem’s scale and speed obfuscate it. Hosting platforms are vast environments constructed through algorithmic logic.\textsuperscript{158} They often operate as “unknowable black box[es],” even to developers.\textsuperscript{159} Civil servants simply lack the knowledge and capacity necessary to scrutinize standards and determinations voluntarily adopted under the guise of goodwill by a platform collective.\textsuperscript{160}

If one ought to be wary of a group of platforms governing together, does it make any more sense to trust them individually to independently self-police and adjudicate speech? In April 2018, Mark Zuckerberg, in an interview with Ezra Klein, described the creation of a Facebook Supreme Court in which “folks at Facebook make the first decision [to take down content] based on . . . community standards that are outlined, and then people get a second


\textsuperscript{156} See Short, supra note 154, at 656–57.

\textsuperscript{157} Freeman seems to implicitly recognize that certain contexts—e.g., informationalism—may be fundamentally incompatible with hybrid regulatory structures or “public/private interdependence.” See Freeman, supra note 155, at 665 (“Public/private regimes may engender doubts insufficiently addressed by the mere existence of agency oversight or the application of familiar procedural controls to private conduct. To be sure, requiring private actors to observe procedures usually demanded only of agencies may in some cases provide minimal accountability.”)

\textsuperscript{158} See generally Christian Sandvig, The Model Did It: Recovering Motives from Machine Learning (forthcoming) (on file with author).

\textsuperscript{159} Id.

\textsuperscript{160} See Cohen, supra note 138, at 179 (“In an era when decision-making is mediated comprehensively by so-called “big data,” regulators seeking to fulfill antidiscrimination mandates must learn to contend with the methods by which regulated decisions are reached—with data and algorithms as instrumentalities for conducting (regulated) activity. In general, the existing regulatory toolkit is poorly adapted for scrutinizing data-driven algorithmic models.”)
opinion.”

Zuckerberg explained that this Facebook Supreme Court would be “made up of independent folks who don’t work for Facebook, who ultimately make the final judgment call on what should be acceptable speech in a community that reflects the social norms and values of people all around the world.”

Though the idea has been slow to develop amid criticism of the fanciful idea of global “social norms,” its driving force — the pervasive problem of hate speech on platforms — seems closely tied to the outbreak of online voter intimidation in recent years.

Facebook has also explored methods to expedite content removal in response to recent high-profile failures with “fake news” and election interference. During 2018, Facebook implemented an “Election War Room”—a Menlo Park conference room converted into a programming battle station to fight “suspicious spikes in spam and hate speech” during the midterm elections. This “War Room” targeted voter suppression efforts through use of threat intelligence, data science, and engineering resources pooled from across Facebook’s enterprise. Over the span of a week, “programmer-soldiers” removed 559 webpages and 251 accounts in the United States that were using fake identities to coordinate fake information campaigns. But, Facebook has not vowed to continue its “War Room” operations. And other platforms that have faced less public backlash seem to be failing in their efforts to police hosted voter intimidation. Although Twitter purged fake accounts and “outlined its efforts to ensure election

---


162 Id.


164 See Lapowsky, supra note 163.


166 Id.

167 Id.

168 Id.
integrity” leading up to the 2018 midterms, it balked at the timely removal of hoax videos and doctored photographs. Even so, Twitter’s performance was, by all accounts, exemplary compared to that of 4chan, which openly abetted vulgar, cruel, and intimidating speech connected to the election. This type of content has become emblematic of 4chan; any hope the public may have for its voluntary removal by site administrators seems badly misplaced.

Assuming, then, that administrative law frameworks—old and new—are inadequate to combat platform-enabled voter intimidation and that more is needed than self-policing by platforms, what other method can be used to protect voters from confusion and undue influence? The last option is to transition to a multi-day election format at the national level. Because the two largest domestic social media providers—Facebook and Twitter—are currently incapable of removing voter intimidation at a rate necessary to nullify effects on voter behavior, the spreading of elections across a longer timeframe would dilute the impact of this speech and give platforms an extended opportunity to take it down. This proposal, though unorthodox,

171 See Appendix 3.1–3.3.
172 With respect to these underperforming platforms, there is always the possibility of amending Section 230 of the Communications Decency Act to allow plaintiffs to bring certain claims directly against them. The legislative solutions discussed above would require conforming amendments to Section 230 in any case. But the edifice of reliance interests built up around Section 230’s “strong protective aura” would spawn herculean political opposition, buttressed by the unusually aligned interests of ISPs and edge service providers. See Kyle Langvardt, Regulating Habit-Forming Technology, 88 Fordham L. Rev. 129, 175 (2019). A more pragmatic alternative might be to simply threaten an amendment—what Daphne Keller describes as “jawboning.” See Keller, supra note 18, at 5. This tactic might lead platforms to augment or initiate self-enforcement efforts, even if actual revision of Section 230 could not be written into law.
would be within the constitutional power of Congress to enact. It would also be practically viable; most states already permit some form of early voting.

Of course, multi-day voting systems are not without their costs. Opening all polling places for additional days drives nationwide increases in staff, security, and facility costs. Meanwhile, the staggering of voting days by district or state can undermine electoral legitimacy. Delays in the announcement of results until all voting has concluded can undermine confidence in the competence of the election’s operation, whereas announcing results daily can affect voting behavior and promote attempts at manipulation. These legitimacy costs, at worst, can exceed the questionable benefits of starving platform-enabled voter intimidation of its viral oxygen. More time will also not change outcomes on bad-actor platforms like 4chan that operate as havens for hate speech. Nor will more time mitigate the effects of insidious, long-lead misinformation campaigns that occur over weeks instead of hours.

In the end, adopting a multi-day election format is a gamble. The monetary, political, and legitimacy costs will be high. The net benefits of providing high-impact platforms more time to eliminate hosted voter intimidation may be marginal, especially considering advances in automated content identification and take-down technologies that these platforms may achieve and implement in the coming years. If the juice is not worth the squeeze, it may be necessary to resort to traditional market pressure and old-fashioned jawboning to drive improved private enforcement of platforms’ rules against voter intimidation in the near term.

VI. CONCLUSION

Platform-enabled voter intimidation is unlikely to subside as social media sites and other third-party forums continue to assume important

---

173 See U.S. CONST. art. II, § 1, cl. 4 (“The Congress may determine the Time of chusing the Electors, and the Day on which they shall give their Votes; which Day shall be the same throughout the United States.”). In fact, the modern system of voting on the first Tuesday after the first Monday in November was initially adopted by Congress in 1845, only because “getting to and from polling places used to be a two-day ordeal, and voting on the weekend or Monday would have meant traveling on the Sabbath.” Juliet Lapidos, Doing Democracy Right, SLATE (Oct. 17, 2008, 5:39 PM), https://slate.com/news-and-politics/2008/10/why-are-other-countries-better-at-conducting-elections-than-we-are.html [https://perma.cc/A84M-3V2V].


176 Id.
communicative and informational functions in modern life. But the existing toolkit of federal voter-intimidation statutes, tailored to the vanishing evil of in-person voter intimidation, is inadequate to protect suffrage. Although Section 11(b) of the Voting Rights Act may be capable of deterring certain long-lead voter intimidation that is attributable to a specific organization, its enforcement mechanisms generally cannot keep pace with individualized threats of violence and anonymized misinformation campaigns that target voters on Election Day.

Furthermore, solutions beyond existing law—traditional administrative law paradigms, hybrid institutional models, and private enforcement—suffer from a combination of constitutional invalidity, logistical impracticability, and unprofitability. The best approach available in the absence of a constitutional revision is to embrace a multi-day election format that blunts the force of “same-day” platform-enabled voter intimidation. Nevertheless, this solution would do little to thwart extended and insidious misinformation campaigns that run the duration of the election cycle, much less to curb the intimidating speech on platforms like 4chan that pridefully abet it. In any case, certain voter intimidation cannot be unseen, even if it can be removed by Facebook and Twitter long before ballots are cast. In the end, the potential costs to electoral legitimacy of a multi-day model may exceed the marginal benefits it provides. In that case, the status quo—waiting for the algorithmic identification systems of leading platforms to pace users’ vast content generation—is the only option.
APPENDIX

1.1

Dillian Billiot

Where's all my Militia brothers? I'm on standby ready to defend our constitution and fight for our freedom. I'll volunteer for this country any day.
#Freedom #America #Constitution #Rights #Milita #Power #WeThePeople #OneNation #WeAreStrongerThanAnyGovernment

1.2

Mc McDonald

The polls opened 12 minutes ago on the east coast. The battle for America has entered a major campaign in the long war of ideologies. The United States either stays a Christian nation or becomes a nation for Islam.
I pray Trump wins as we veterans and militia are ready to take our country back. You civilians either follow us or get out of the way. Your only other option is to be a quantified target.

1.3

Jack Daniel Bridges

Keep em locked and loaded folks. If Tensing is found not guilty
if Hillary wins. LOCKED AND LOADED!

1.4

Missie Hendricks

THEY RIGGED THE PRIMARIES & PEOPLE TOOK IT. IF THEY RIG TODAY'S ELECTION, IT'S WAR.
1.5

It's LOCK AND LOAD if Donald J. TRUMP isn't PRESIDENT.

1 share

1.6

If Killary is elected president... Don't call me looking for help when shit hit the fan... locked and loaded my friends... till the end

13

2.1

I've just got word that an illegal trying to vote has been arrested.

ICE is watching voting places closely! Thank you ICE!

More to come!
5:18 PM - 27 Oct 2016

2.2
What if we lose?
Then we'll meet IRL.

BuzzFeed News reporter Joe Bernstein is watching the discussion threads on noted troll message boards 4chan and 8chan. He's spotted two plans being hatched to spread false information. This one involves hoaxing Spanish-language media:
Our third panel focused on our election infrastructure and voting procedures. The Help America Vote Act introduced computerized voting, in the form of Direct Recording Electronic (DREs), Ballot Marking Devices (BMDs) and computerized scanners, to polling places across the United States. And on a more local level, counties and states manage everything from ballot design to result reporting with computers, often in ways that are exposed to the Internet. Modern elections are now therefore at least partly dependent on the integrity of the software and hardware on which they run, and this panel looked to discover the ways in which those dependencies created vulnerabilities and what solutions for those vulnerabilities there might be.

Matt Blaze, Professor of Law and Computer Science at Georgetown University, led this panel. He was joined on stage by Andrew Appel of Princeton University, Barbara Simons of Verified Voting, Philip Stark of the University of California, Berkeley, and Kim Zetter, an award-winning freelance journalist whose work on electronic voting machines has been featured in Wired magazine and the New York Times. C-SPAN has a recording of this panel, available at https://www.c-span.org/video/?469098-4/georgetown-law-discussion-election-integrity.

We are pleased to present articles by Mr. Blaze, Mr. Appel and Mr. Stark, and Ms. Simons in the pages that follow. This panel was the most technical of the day, and we hope that you find these articles accessible and informative.
ELECTION INTEGRITY AND TECHNOLOGY: VULNERABILITIES AND SOLUTIONS

Matt Blaze*

CITE AS: 4 GEO. L. TECH. REV. 505 (2020)

TABLE OF CONTENTS

I. ELECTIONS AND SOFTWARE SECURITY ....................................................... 505
   A. Election Processes, Software, and Hardware ........................................ 507
   B. Software and Election Security ............................................................. 509
II. CURRENT ELECTRONIC VOTING SYSTEMS HAVE PROVEN
    VULNERABLE TO A RANGE OF KNOWN, EXPLOITABLE SECURITY FLAWS .... 510
    A. Risks in Various Election Components ............................................. 510
       1. Election Management IT Systems .................................................. 511
       2. Electronic Poll Books ...................................................................... 512
       3. Optical Scan Ballot Readers ............................................................. 512
       4. Ballot Marking Devices ..................................................................... 513
       5. Direct Recording Electronic (DRE) Voting Machines ....................... 513
    B. The 2007 California and Ohio Studies .............................................. 515
    C. The DEFCON Voting Village Exercise ............................................... 516
III. U.S. ELECTION SYSTEMS ARE NOT ENGINEERED TO RESIST
     NATIONAL ADVERSARIES ...................................................................... 518
IV. MOVING FORWARD .................................................................................... 520

I. ELECTIONS AND SOFTWARE SECURITY

A consequence of our federalist system and tradition is that even though U.S. elections are organized around a national hierarchy, they are

*McDevitt Professor of Law and Computer Science, Georgetown University Law Center and Georgetown University Department of Computer Science. This Article is taken from my testimony before the House of Representatives on two separate occasions. On November 19, 2019, I delivered testimony to the Committee on Homeland Security in a hearing titled Defending Against Election Interference, and on January 9, 2020, I delivered additional testimony to the Committee on House Administration in a hearing titled 2020 Election Security-Perspectives from Voting System Vendors and Experts. This testimony has been lightly edited for inclusion in this issue of GLTR.
executed in a highly decentralized manner, with each state responsible for setting its own standards and procedures for registering voters, casting ballots, and counting votes. The federal government sets broad standards for such issues as accessibility but has historically been largely uninvolved in day-to-day election operations. In most states, the majority of election management functions are delegated to local county and town governments, which are responsible for registering voters, procuring voting equipment, creating ballots, setting up and managing local polling places, counting votes, and reporting the results of each contest. Thousands of individual local election offices shoulder the burden of managing and securing the voting process for most of the American electorate.

Consequently, elections in the United States are among the most operationally and logistically complex in the world, and information technology plays an essential role in overcoming that complexity. Many jurisdictions have large numbers of geographically dispersed voters, and most elections involve multiple ballot contests and referenda. Computers and software enable our election process by managing voter registration records, defining ballots, provisioning voting machines, tallying and reporting results, and controlling electronic voting machines used at polling places.¹

Computers became ingrained into our election system with the passage of the Help America Vote Act (HAVA) in 2002.² The Act accelerated the computerization of voting systems, particularly with respect to the ways in which voters cast their ballots at local polling stations. HAVA provided funds for states to replace precinct voting equipment with “accessible” technology. As implemented, however, some of this new technology has had the unfortunate unintended consequence of increasing, rather than decreasing, the risk of our elections being compromised by malicious actors.

The integrity and security of our elections are inexorably tied to the integrity and security of the computers and software that we rely on for these many functions, and election security must account for sophisticated adversaries, ballot secrecy, fair access to the polls, and accurate reporting of results. All of these concerns make secure election management one of the most formidable—and potentially fragile—information technology problems in government.

---

¹ A typical election administration office is much like any modern enterprise, with local computer networks tying together desktop computers, printers, servers, and Internet access. In 2016, this increasing connectivity served as a critical avenue for what U.S. intelligence agencies have identified as attacks by Russian military intelligence. S. SELECT COMM. INTELLIGENCE, REPORT ON RUSSIAN ACTIVE MEASURES IN THE 2016 US ELECTION, VOL 1., S. REP. NO. 116-XX (2019).

In this Paper, I will give an overview of the technical security risks facing elections in the United States today, with an emphasis on vulnerabilities inherent in electronic voting machines as well as the exposure of our election infrastructure to disruption by domestic as well as national security adversaries. It is, by necessity, incomplete and narrow. But I do hope that this Paper highlights some of the ways that technology has enabled the election process while raising awareness about that technology’s vulnerabilities.

An especially valuable resource, with comprehensive discussion and recommendations is the recent National Academies Securing the Vote consensus study report.

A. Election Processes, Software, and Hardware

A typical county election office today depends on computerized systems and software for virtually every aspect of registering voters and conducting elections. Generally, an election office workflow will include at least the following pre- and post-election functions:

Voter registration: The ongoing maintenance of an authoritative database of registered voters in the jurisdiction, including the precinct-by-precinct “poll books” of voters (which might be on paper or in electronic form) that are used to check in voters at precinct polling stations.

Ballot definition: The pre-election process of creating data files that list the various contests, candidates, and rules (e.g., number of permitted choices per race) that will appear on the ballot. The ballot definition is used to print paper ballots, to define what is displayed on touchscreen voting terminals, and to control the vote tallying and reporting software. Local races (such as school boards) may sometimes require that different ballot definitions be created for different precincts within a county in any given election.

---

3 This short paper is focused narrowly on technical vulnerabilities and threats specific to election administration and the voting processes itself. This paper does not attempt to cover other serious threats to elections even though they may leverage modern technology (such as, for example, disinformation campaigns that exploit digital media).


5 The precise nature of the systems used and how they interact with one another will vary somewhat depending on the vendors from which the systems were purchased and the practices of the local jurisdiction.
Voting machine provisioning: The pre-election process of configuring the individual precinct voting machines for an election. This typically includes resetting internal memory and loading the appropriate ballot definition for each precinct. Depending on the model of voting machine, provisioning typically involves using a computer to write removable memory cards that are installed in each machine.

Absentee and early voting ballot processing: The process of reading and tabulating ballots received by mail and from early voting polling places. Mail votes are typically processed in bulk by high-volume optical scan ballot reading equipment.

Tallying and reporting: The post-election process of tabulating the results for each race received from each precinct and reporting the overall election outcomes. This process typically involves using a computer to read memory card media retrieved from precinct voting machines.

Each of these “back end” functions employ specialized election management software, usually running on standard desktop and laptop computers. Depending on the size and practices of the county, the same computers may be used for more than one function (e.g., the ballot definition computer might also serve as the tallying and reporting computer). These computers are typically off-the-shelf desktop machines running a standard operating system (such as Microsoft Windows), often equipped with electronic mail and web browser software along with the specialized voting software. Election office computers are typically connected to one another via a wired or wireless local area network, which may have a direct or indirect connection (sometimes via a firewall) to the Internet.

In some jurisdictions, some or all election management functions (most typically those concerned with voter registration databases and ballot definition) may be outsourced by a county or state to an election services contractor. These contractors provide jurisdictions with specialized assistance with such tasks as creating ballots in the correct format, managing voter registration databases, creating precinct poll books, and maintaining voting machines. The degree to which jurisdictions rely on outside contractors varies widely across the nation.

Much of the voting equipment used at precincts is computerized as well, although it is generally packaged in specialized hardware. This equipment includes:

Direct Recording Electronic (DRE) Voting Machines: DRE machines are special-purpose computers that display ballot choices to the voter
Optical Scan Ballot Readers: Optical scan ballot readers are specialized computers that read voter-marked paper ballots. The ballot is read according to the ballot definition configuration (typically on removable memory media), and a tally is maintained in memory (also typically on removable media). The machine also captures the scanned ballots and stores them in a mechanically secured ballot box.

Ballot Marking Devices (BMDs): Ballot marking devices are an assistive technology used in optical scan systems to allow visually or mobility impaired voters to create ballots for subsequent scanning. BMDs are similar in appearance to DRE machines in that they display (or read aloud) the ballot electronically, based on a ballot definition configuration, and accept voter choices for each race. However, instead of recording those choices in computer memory as DREs do, BMDs print a marked paper ballot that can then be submitted through an optical scan ballot reader.

Electronic Poll Books: These devices are typically tablet-style computers that contain an authoritative copy of the database of registered voters at each precinct. Electronic poll books are not used directly by voters, but rather by precinct poll workers as voters are checked in at their polling place. They are not used in all jurisdictions.

B. Software and Election Security

Securing complex software systems is notoriously difficult, and those that perform the various voting and election management functions described above are no exception. There are several potential avenues of technical

---

6 Some models of DRE can be equipped with a Voter Verified Paper Audit Trail (VVPAT) option in which the voters’ selections are printed on a paper tape roll that is visible to the voter. VVPATs can assist with determining the voter’s intent during a recount, but their efficacy depends on each voter’s diligence in confirming that their choices are correctly recorded on the paper tape before they leave the voting booth. Research consistently suggests that, in practice, very few voters successfully perform this confirmation step. See, e.g., Andrew W. Appel & Philip B. Stark, Evidence-Based Elections: Create a Meaningful Paper Trail, Then Audit, 4 GEO. L. TECH. REV. 523 (2020).

7 The fact that software systems can be, and often are, vulnerable to attack is not unique to election systems, of course. Serious data breaches are literally daily events across the public
vulnerability in such systems. Common software “bugs” often introduce vulnerabilities that can be exploited by an adversary to silently compromise the integrity of data or make unauthorized (and difficult to detect) changes to the behavior of systems. Configuration and system management errors (such as the use of vulnerable out-of-date platforms and weak passwords) can further expose systems to tampering. Computer networks (which are not generally used by precinct voting machines themselves but are commonly connected to back end systems in election offices) compound these risks by introducing the possibility of remote attack over the Internet.

The integrity of the vote today thus increasingly depends on the integrity of the software systems—running on voting machines and on county election office networks—over which elections are administered and vote tallies conducted. Any security weakness in any component of any of these systems can serve as a “weak link” that can allow a malicious actor to disrupt election operations, alter tally results, or disenfranchise voters.

And in many electronic voting systems used today, a successful attack that exploits a software flaw might leave behind little or no forensic evidence. This can make it effectively impossible to determine the true outcome of an election or even that a compromise has occurred.

Unfortunately, these risks are not merely hypothetical or speculative. Many of the software and hardware technologies that support US elections today have been shown to suffer from serious and easily exploitable security vulnerabilities that could be used by an adversary—insider or outsider—to alter vote tallies or cast doubt on the integrity of election results.

II. CURRENT ELECTRONIC VOTING SYSTEMS HAVE PROVEN VULNERABLE TO A RANGE OF KNOWN, EXPLOITABLE SECURITY FLAWS

A. Risks in Various Election Components

Security concerns about computerized voting systems have been raised from almost the moment such systems were first proposed. Most of these concerns have focused on electronic voting equipment used at polling stations. However, the “back end” election management software used to manage voter registration, provision voting machines, and tally voted are at least equally critical to the integrity of the vote.

To be clear, all current electronic voting technology can and does suffer from security vulnerabilities. Whether successfully exploiting these vulnerabilities has significant consequences, however, depends on the

and private sectors, and cybersecurity is widely recognized to be a serious law enforcement and national security problem. To the extent that elections depend on software or are administered by networked computing systems, they are subject to all the same risks.
particular class of device and whether the technology permits effective post-election auditing to validate or recover accurate election results and detect anomalies.

1. *Election Management IT Systems*

As noted in the previous section, local jurisdictions rely on computers for almost every aspect of election administration. Official information for voters is distributed on public-facing websites. Voter registration records, used on election day to determine who is permitted to vote, are maintained in computerized databases. Ballot forms are created and edited on computers. Absentee ballot mailings are managed by computer. Preliminary and official election results are maintained and disseminated by computer. Specialized “Election Management” software (generally provided by the vendor of the jurisdiction’s voting equipment) is used to configure ballots and read results from precinct voting machines.

In most cases, the computers used for election administration employ the same hardware, operating systems, and networking platforms employed by other enterprises, all of which may be directly or indirectly connected to the Internet. Election management systems are therefore exposed to the same risks of compromise seen in other private and public sector domains; the malicious actors that perpetrate the all-too-common data breaches of our regular online life can reach our election systems as well.

Many jurisdictions outsource some of their election management tasks to outside vendors or contractors: this practice amplifies the exposure of election infrastructure to external tampering.

Disruption or compromise of any local election administration functions can have grave (and often non-recoverable) consequences for election integrity. Adversaries can compromise voter registration databases to cause long lines at polling places (forcing large numbers of voters to cast provisional ballots) and to selectively disenfranchise voters to favor particular candidates. Voting machines with incorrect ballot definitions can prevent correct ballots from being cast. Errors in in unofficial or final tallies can cast doubt on the legitimacy of entire elections. In some cases, successful attacks may not be discovered until long after polls have closed or may never be discovered at all.

The IT and security administration of election management computers varies widely from jurisdiction to jurisdiction. In the best cases, there may be a full-time staff devoted to securing and managing election computers and

8 Most election jurisdictions, like other enterprises, employ “firewalls” between their internal networks and the public Internet. However, firewalls are not by themselves a complete or sufficient defense against remote attack.
networks. In a more typical case, computer security is relegated to the general county IT staff, who may have limited resources relative to the threat. In all cases, however, a local county’s strongest cybersecurity resources are of only limited value against a foreign state adversary.

Local election management computers and networks are especially attractive targets for foreign tampering and interference. They can often be attacked remotely, without the need for physical presence in the targeted jurisdiction, and successful attacks may be rewarded with partial or complete control over a county’s voter registration databases, voting machine configuration, and results reporting infrastructure.

2. Electronic Poll Books

A poll book, or electoral roll, is a list of eligible and registered voters, and some jurisdictions use electronic poll books to perform the initial voter “check in” function at polling places on election day. They must, by nature of their function, have reliable access to an authoritative list of the voters registered to vote at each polling place. This may be accomplished either with an internal copy of the voter registration database or by online remote access to a central computer. In either configuration, electronic poll books perform an essential election function and must be reliably secured against tampering. If poll books are unavailable or if their databases are corrupted, voters will not be able to cast ballots (except by provisional ballot, to the extent that is a viable option).

Electronic poll books have received far less scrutiny than other precinct voting equipment, but these poll books are subject to all the same risks and attack vectors as other electronic devices. In many jurisdictions, they are largely unregulated and require little or no outside certification or audit.

3. Optical Scan Ballot Readers

Optical scan ballot readers are specialized computers that scan and retain printed ballots and record the tally of votes cast in each race on electronic storage media. They depend on the integrity of their software and hardware for their ability to correctly interpret ballots and to correctly record votes. They are exposed to physical access by poll workers, and, in many cases, individual voters.

Ballot scanners can be compromised in a number of practical ways, any one of which can compromise the recorded vote tally. However, because they retain the physical paper ballots marked by voters, it is possible to recover from such a compromise, if detected. A technique called “risk-limiting audits”
can reliably detect and recover from defective or compromised ballot scanners and is discussed in the sections that follow.

4. **Ballot Marking Devices**

Originally, Ballot Marking Devices (BMDs) were conceived of narrowly, as an assistive technology to assist voters with disabilities in marking optical scan paper ballots.9 (However, certain recent voting products greatly expand the use of BMD technology by integrating a BMD into the voting process for all voters, whether they require assistive technology or not.)

BMD-based voting systems are controversial because, by virtue of their design, their behavior cannot be effectively audited except by individual voters carefully verifying their machine-printed ballots before they are cast. A maliciously compromised BMD could subtly mismark candidate selections on ballots in a way that might not be noticed by voters and that could undetectably alter election outcomes. Furthermore, if BMDs fail or must be rebooted at a polling place, there may be no alternative method for voters to create marked ballots, making BMDs a potential bottleneck or single point of failure on election day.

As a relatively new technology, BMD-based systems have not yet been widely examined by independent researchers and have been largely absent from practical election security research studies. However, even with relatively little scrutiny, exploitable weaknesses and usability flaws have been found in these systems. This underscores the need for more comprehensive studies and for caution before these systems are purchased by local jurisdictions or widely deployed.

5. **Direct Recording Electronic (DRE) Voting Machines**

From a security perspective, by far the most problematic and risky class of electronic voting systems are those that employ Direct Recording-Electronic (DRE) machines.10 DRE machines are special purpose computers programmed to present the ballot to the voter and record the voter’s choices on an internal digital medium such as a memory card. At the end of the election day, the memory card containing the vote tallies for each race is generally removed or electronically read from the machine and delivered to the county

---

9 This update had the effect of bringing such systems into compliance with Help America Vote Act (HAVA) requirements for accessible voting. See Help America Vote Act of 2002, supra note 2.

10 DRE machines are sometimes informally called “touchscreen” voting machines, although not all DRE models use actual touchscreen displays. Moreover, not all election devices that employ touchscreens DREs.
election office, where the tallies from each precinct are recorded by the county tallying software.

The design of DREs makes them inherently difficult to secure and also makes it especially imperative that they be secure. This necessity exists because the accuracy and integrity of the recorded vote tally depends completely on the correctness and security of the machine’s hardware, software, and data. Every aspect of a DRE’s behavior, from the ballot displayed to the voter to the recording and reporting of votes, is under control of the DRE hardware and software. Any security vulnerability in this hardware or software, or any ability for an attacker to alter (or re-load new and maliciously behaving) software on the machine, not only has the potential to alter the vote tally, but can make it impossible to conduct a meaningful recount (or even to detect that an attack has occurred). If a DRE is compromised at any time before or during an election, any votes cast on it are irreparably compromised as well.12

---

11 An incorrect (or maliciously altered) DRE ballot definition can make it impossible to determine the true results of an election even without any malicious software exploitation on another part of the machine. For example, in a 2017 York County, Pennsylvania, local election, a programming error on the county’s DRE machines allowed voters to cast two votes for candidates in contests where there was more than one winner. David Weissman, York County Details Lack of Internal Controls in Post-Election Report to State, YOREAL DAILY RECORD (Dec. 7, 2017, 5:02 PM), https://www.yorkdispatch.com/story/news/local/2017/12/07/york-county-details-lack-internal-controls-post-election-report-state/931308001. Because the DRE machines filled in and printed the voters’ ballots, the county’s solution was to identify how many times there was an “overvote,” in which a candidate received two votes from the same voter. David Weissman, York County Officials Say Overvotes Didn’t Affect Election Results, But Numbers Tell a Different Story, YORK DAILY RECORD (Nov. 20, 2017, 3:01 PM), https://www.yorkdispatch.com/story/news/local/2017/11/20/york-county-officials-say-overvotes-didnt-affect-election-results-but-numbers-tell-different-story/881496001. Nine races were affected, and even though the candidate chose not to challenge the results, just nine votes decided one race in which there were thirty-two erroneous “overvotes.” Jason Addy, York County Commissioners Certify Unchallenged Election Results, YORK DAILY RECORD (Nov. 27, 2017, 5:14 PM), https://www.yorkdispatch.com/story/news/politics/elections/2017/11/27/york-county-commissioners-certify-unchallenged-election-results/898381001.

12 In contrast, systems that do not rely on a machine to display and register a vote, such as paper-based systems, are more robust against such errors. For example, the 2000 general election in Bernalillo County, New Mexico, experienced a programming mistake in its vote counting software, but the county was able to correct the error without a new election by recounting the paper ballots, which had not been produced by a vote casting machine. See Bryan Gruley & Chip Cummins, Election Day Became a Nightmare, As Usual, for Bernalillo County, WALL ST. J. (Dec. 15, 2000, 11:59 PM), https://www.wsj.com/articles/SB976838091124686673.
DRE-based systems introduce several avenues for attack that are generally not present (or are not as security-critical) in other voting technologies:

- Alteration or deletion of vote tallies stored in internal memory or removable media;
- Alteration or deletion of ballot definition parameters displayed to voters; and
- Alteration or deletion of electronic log files used for post-election audits and detecting unauthorized tampering.

Attacks might be carried out in any of several ways, each of which must be reliably defended against by the DRE hardware and software:

- Direct tampering with data files stored on memory cards or accessible through external interface ports;
- Surreptitious replacement of the certified software running on the device with a maliciously altered version; and
- Exploitation of a pre-existing vulnerability in the certified software.

Successfully exploiting just one of these avenues of attack can be sufficient to undetectably compromise an election. Because of the DRE machines’ design, users must not only guarantee that the machines’ hardware is secure against unauthorized tampering but also that the software running on the machines does not suffer from any vulnerabilities that could be exploited by a malicious actor. This makes the security requirements for DREs more stringent—and more easily defeated—than for any other currently deployed election technology.

Unfortunately, the DRE-based systems purchased by (and still used in) various states under HAVA have repeatedly been found to suffer from exactly these kinds of exploitable hardware and software vulnerabilities.

B. The 2007 California and Ohio Studies

To date, the most extensive independent studies of the security of electronic voting systems were commissioned in 2007 by the Secretaries of State of California and Ohio. Expert review teams were given access to the voting machine hardware and software source code of every system certified for use in those states. The systems used in California and Ohio were also certified for use in most of the rest of the country, so these studies effectively covered a large fraction of available electronic voting equipment and software. The author led the teams that reviewed Sequoia products (for the state of
California) and ES&SS products (for the state of Ohio); other teams in these studies reviewed Diebold/Premier and Hart InterCivic products.\(^{13}\)

In both studies, every team found and reported serious, exploitable vulnerabilities in *almost every component* examined. In most cases, these vulnerabilities could be exploited by a single individual, who would need no more access than an ordinary poll worker or voter to carry out effective attacks. Such an attacker would be able to alter vote tallies, load malicious software, or erase audit logs. Some of the vulnerabilities found were the consequence of software bugs, while others were caused by fundamental architectural properties of the system architecture and design. In some cases, compromise of a single system component (such as a precinct voting machine) was sufficient to compromise not just the vote tally on that machine, but to compromise the entire county back end system.

In response, California and Ohio ordered some equipment decertified and some election-day procedures modified. However, all the vulnerable equipment and software remained certified for use in at least some other states.

Some equipment vendors and local voting officials claimed at the time that the findings of the California and Ohio studies were irrelevant or overstated, that any problems identified could be easily fixed, and that it would be difficult or impossible for anyone but an expert with extensive experience and access to privileged information (such as source code) to exploit vulnerabilities in practice. However, as exercises such as the DEFCON Voting Village (described below) have demonstrated, not only do these systems remain vulnerable, but they can be readily exploited by people with no more than ordinary, undergraduate-level computer science experience and expertise, and without access to any secret or proprietary information.

C. The DEFCON Voting Village Exercise

The DEFCON conference is one of the world’s largest and best-known computer security “hacker” conferences. Last year’s DEFCON was held August 8–10, 2019, in Las Vegas, Nevada, and drew more than 25,000 participants from around the world. DEFCON participants have broad interest

\(^{13}\) *See generally Deborah Bowen, Cal. Secretary of St., Top-to-Bottom Review (2007)* [https://www.sos.ca.gov/elections/ovsta/frequently-requested-information/top-bottom-review](https://www.sos.ca.gov/elections/ovsta/frequently-requested-information/top-bottom-review) [https://perma.cc/FC6D-3RX4]; EVEREST: EVALUATION AND VALIDATION OF ELECTION-RELATED EQUIPMENT, STANDARDS & TESTING (2007), [https://www.cise.ufl.edu/~butler/pubs/everest.pdf](https://www.cise.ufl.edu/~butler/pubs/everest.pdf) [https://perma.cc/RBU9-MBH3] (prepared by teams from Pennsylvania State University, the University of Pennsylvania, and WebWise Security, Inc. as part of the EVEREST voting systems analysis project initiated by the Secretary of State of Ohio).
in technology, and include security researchers from industry, government, and academia, as well as individual hobbyists.

For the last three years, DEFCON has featured a voting machine hacking village (the Voting Village) to give participants an opportunity to examine and get hands-on experience with the security technology used in U.S. elections, including voting machines, voter registration databases, and election office networks. The author is one of the organizers of the Voting Village.\textsuperscript{14}

The voting machines available in the Voting Village included a variety of DRE, optical scan readers, ballot marking devices, and electronic poll books from a range of commercial vendors. We acquired (from the surplus market) and made available to participants a sampling of different pieces of election hardware, including both DRE and optical scan voting machines as well as “poll book” devices used by used by precinct workers to verify and check in voters at polling places. Every model machine currently at the Voting Village is still certified for use in U.S. elections in at least one jurisdiction today.

The DEFCON Voting Village is not intended to be a formal security assessment or test, but rather an opportunity for a general audience of technologists to examine election equipment and systems. However, participants are encouraged to critically examine and probe the equipment and software for vulnerabilities, and to seek practical ways to compromise security mechanisms. No proprietary information or computer source code is made available.

The results of the Voting Village are summarized each year in detail in a report.\textsuperscript{15} It is notable that participants, who overwhelmingly do not have any previous special expertise in voting machines or access to any proprietary information about them, have been very quickly able to find ways to compromise every piece of equipment in the Village by the end of the weekend. Depending on the individual model of machine, participants have found ways to load malicious software, gain access to administrator passwords, compromise recorded votes and audit logs, or cause equipment to fail. In most cases, these attacks could be carried out from the ordinary interfaces that are exposed to voters and precinct poll workers.


\textsuperscript{15} BLAZE ET AL., supra note 14.
The ease with which participants compromise equipment in the Voting Village should be regarded as at once alarming and yet also unsurprising. It is alarming because the very same equipment is in use in polling places around the United States, relied on for the integrity of real elections. But it is also ultimately unsurprising. Versions of many of the machines at DEFCON had been examined in the 2007 studies and found to suffer from basic, exploitable security vulnerabilities. It should not come as any surprise that, given access and motivation, people of ordinary skill in computer security would be able to replicate and expand on these results. It is, in fact, precisely what the previous studies of these devices warned would happen.

In summary, the DEFCON Voting Village demonstrates that much of the voting technology used in the United States remains vulnerable not just to hypothetical expert attack in a laboratory environment, but also to practical analysis, manipulation and exploitation by non-specialists with only very modest resources.

III. U.S. ELECTION SYSTEMS ARE NOT ENGINEERED TO RESIST NATIONAL ADVERSARIES

The traditional “threat model” against which electronic voting systems have been evaluated has been largely focused on resisting traditional election fraud, in which domestic conspirators, perhaps assisted by corrupt poll workers or election officials, attempt to “rig” an election to favor a preferred candidate in a local, state, or national contest. Fraud might be accomplished by altering votes, adding favorable votes, deleting unfavorable votes, or otherwise compromising the security mechanisms that protect the ballot and tally. While virtually every study of electronic voting technology has raised questions about the ability of current systems to resist serious efforts at fraud, traditional election fraud is not the only kind of threat, or even the most serious threat, that voting systems must resist today.

Electronic voting systems must resist not only fraud from corrupt candidates and supporters, but also election disruption from hostile foreign adversaries. This is a much more formidable threat, and one that current systems are far less equipped to resist.

The most obvious difference between traditional election fraud by corrupt domestic actors and disruption by hostile state actors is the expected resources and capabilities available to each. The intelligence services of even small nations can marshal far greater financial, technical, and operational resources than would be available to even highly sophisticated criminal conspiracies. For example, intelligence services can feasibly conduct advance operations against the voting system supply chain. In such operations, the aim might be to obtain confidential source code or to secure surreptitious access to
equipment before it is even shipped to local election officials. Hostile intelligence services can exploit information and other assets developed broadly over extended periods of time, often starting well before any specific operation or attack has been planned.

But their greater resources are not the most important way that hostile state actors can be a more formidable threat than corrupt candidates or poll workers. They also enjoy easier goals. The aim of traditional “retail” election fraud is to tilt the outcome in favor of a particular candidate. That is, to succeed, the attacker must generally alter the reported vote count or add, change, or delete votes. But a hostile state actor—via an intelligence service such as Russia's GRU—might be satisfied with merely disrupting an election or calling into question the legitimacy of the official outcome. With election systems so heavily dependent on demonstrably insecure software and voting equipment, this kind of disruption could be comparatively simple to accomplish, even at a national scale.

A hostile state actor who can compromise even a handful of county networks might not need to alter any actual votes to create widespread uncertainty about an election outcome’s legitimacy. It may be sufficient to simply plant suspicious (and detectable) malicious software on a few voting machines or election management computers, create some suspicious audit logs, delete registered voters from the rolls, or add some obviously spurious names to the voter rolls. If the preferred candidate wins, they can simply do nothing. If the “wrong” candidate wins, however, they could covertly reveal evidence that county election systems had been compromised, creating public doubt about whether the election had been “rigged.” This could easily impair the ability of the true winner to effectively govern, at least for a period of time.

Electronic voting machines and vote tallies are not the only potential targets for such attacks. Of particular concern are also the “back end” systems that process voter registration, ballot definition, and other election management tasks. Compromising any of these systems (which are often connected, directly or indirectly, to the Internet and therefore potentially remotely accessible) can be sufficient to disrupt an election while the polls are open or cast doubt on the legitimacy of the reported result. Decentralized election operations, which are managed by thousands of individual local offices with various levels of resources throughout the nation, is sometimes cited as a strength of our electoral process. However, this decentralization can be turned to the adversary’s advantage. An attacker can choose arbitrarily from

---

16 Or, in a best-case scenario for any potential adversary, these malicious actors could use their previously arranged access to restore the compromised networks to their original states, erasing any evidence of compromise.
among whatever counties have the weakest systems—those with the least secure software or most poorly defended networks and procedures—to target.

It is beyond the scope of this paper to speculate on specific intrusions that occurred against state and local election management systems in the 2016 U.S. general election, much of which remains classified or under investigation. It has been reported that voter registration management systems in at least several states were targeted for exploitation and access. It is unclear whether voting machines or tallying systems were also targeted. However, targeting and exploiting such systems would have been well within the capability of any major rival intelligence service.\(^\text{17}\)

In summary, the architecture of many current electronic voting systems, especially those that employ DRE voting machines, makes disruption attacks an attractive option for our foreign adversaries—and an especially difficult one to effectively defend against. These systems can give hostile actors interested in disruption an even easier task than that facing corrupt candidates seeking to steal even a small local office. And the consequences of election disruption strike at the very heart of our national democracy.

**IV. MOVING FORWARD**

It is perhaps tempting to conclude pessimistically that election technology in the United States is fatally flawed, leaving our nation irreparably vulnerable to election fraud and foreign meddling. But while it is true that the current situation exposes us to significant risk, it is by no means hopeless or beyond repair. Relatively simple, and available, technologies can be deployed that render our elections significantly more robust in the face of attack.

While electronic voting machines do indeed suffer demonstrably fundamental weaknesses, some electronic voting technologies are significantly more resilient in the face of compromise than others. The most important feature required is that there be a reliable record of each voter’s true ballot selections that can be used as the basis for a post-election audit to detect and recover from failure or compromise of the software or hardware.

Among currently available, HAVA-compliant voting products, the only systems that meet this requirement are those that employ optical scan paper ballot technology. In such systems, the voter fills out a machine-readable paper ballot form (possibly with the aid of an assistive ballot marking device for language-, visually- and mobility-impaired voters) that is then deposited into a ballot scanning device that reads the ballot choices, maintains an electronic tally, and retains and secures the marked paper ballots for

subsequent audit. After the polls close, the electronic tally records are read from each ballot scanner and preliminary results calculated.

The paper records of votes that precinct-counted optical-scan systems provide are a necessary, but not by themselves sufficient, safeguard against software attacks or irregularities. As noted above, even non-DRE systems can suffer from flaws and exploitable vulnerabilities in the voting machine and back end software. The second essential safeguard is a systematic and reliable process for detecting whether the software has reported incorrect results and to recover the true results if so.

The most reliable and well-understood method to achieve this is through an approach called risk-limiting audits. In a risk limiting audit, a statistically rigorous method is used to select a randomized sample of ballots, which are manually checked by hand and compared with their electronic interpretation. (This must be done for every contest, not just those with close results that might otherwise call for a traditional “recount”.) If discrepancies are discovered between the manual and electronic tallies, additional manual checks are conducted. The effect of risk-limiting audits is not to eliminate software vulnerabilities, but to ensure that the integrity of the election outcome does not depend on the herculean task of securing every software component in the system. This important property is called strong software independence.

It is worth emphasizing that risk-limiting audits are only meaningful if there is a reliable, human-readable artifact of the voters’ true selections, such as paper ballots that have been directly marked by the voter.

Optical scan paper ballots and risk-limiting audits comprise a critical—and readily deployable—safeguard against both traditional election fraud and national security threats. Taken together, they permit us to more safely enjoy the benefits of computerized election management, without introducing significant new costs or requiring the development of speculative new technology. The technology required for this is available today, from multiple vendors, and is already in use in many states. In jurisdictions that already use optical scan ballots, implementing effective risk-limiting audits is entirely a procedural matter. In those that do not, it will also require the investment in new precinct voting equipment.

As important as paper ballots and risk-limiting audits are, however, they are not panaceas that solve every threat to our elections. It is equally

---


critical that the state and county computer infrastructure used for election management and voter registration be vigilantly protected against compromise. As we saw in 2016, hostile actors—whether foreign or domestic—might attempt to breach not just voting machines, but also back end election management systems and voter registration database systems, which are often exposed to remote attack over the Internet.

It is no exaggeration to observe that state and local election officials serve on the front lines of our national cybersecurity defense. They must be given sufficient resources, infrastructure, information, and training to help them effectively defend their systems against an increasingly sophisticated—and increasingly aggressive—threat environment. It is notable that the budgets for election administration often must compete for resources with essential local services such as fire protection and road maintenance. Election management represents only a miniscule fraction of the total national spending on political campaigns. Additional investment here will pay significant dividends for our security.

By analogy, we do not make the county sheriff responsible for defending against ground invasions by foreign military forces. Yet that is precisely the role into which we have placed local county IT administrations in defending election infrastructure against electronic attacks. Without significant national-level support, they are doomed to fail.

Simply put, much of our election infrastructure remains vulnerable to practical attack, with threats that range from traditional election tampering in local races to large-scale disruption by national adversaries. We should take no comfort if such attacks have not yet been widely detected. At best, it is only because, for whatever reason, serious attempts have not yet been made. Given the potential rewards to our adversaries, it is only a matter of time before they will.

National-level investment in safeguards such as those described above serve our democracy in critically important ways. They can provide a significant improvement to election security, both in our ability to resist attack and in our ability to recover from attacks when they occur. Perhaps most importantly, they provide meaningful assurance to voters that their ballots truly count and that their elected officials are governing truly legitimately.
EVIDENCE-BASED ELECTIONS: CREATE A MEANINGFUL PAPER TRAIL, THENAudit

Andrew W. Appel* & Philip B. Stark**

CITE AS: 4 GEO. L. TECH. REV. 523 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ............................................................................................................. 523
II. VOTER-VERIFIED PAPER BALLOTS ................................................................. 525
    A. Hand-Marked Paper Ballots (Optical Scan) .............................................. 525
    B. Direct-Recording Electronic (DRE) Machines ........................................ 526
    C. Voter-Verifiable Paper Audit Trail (VVPAT) ............................................. 526
    D. Ballot-Marking Devices (BMDs) ................................................................. 527
    E. All-In-One BMDs .......................................................................................... 528
    F. Internet Voting ............................................................................................... 529
    G. Software Independence, Contestability, Defensibility . . . . . . . . . . . . . . . . . . . . . . . . . . 529
III. RISK-LIMITING AUDITS .................................................................................. 530
IV. COMPLIANCE AUDITS ..................................................................................... 532
V. EFFICIENT RISK-LIMITING AUDITS ....................................................... 534
VI. RESOURCES FOR RISK-LIMITING AUDITS ........................................... 536
    A. Audit the Digital Images? ............................................................................. 537
VII. PRINCIPLES FOR ELECTION INTEGRITY LEGISLATION ...................... 537
VIII. CONCLUSIONS .................................................................................................... 540

I. INTRODUCTION

There is no perfect, infallible way to count votes. All methods—including optical scan, touchscreen, and hand counting—are subject to errors, procedural lapses, and deliberate manipulation. Almost all U.S. jurisdictions count their votes using computer-based technology, such as touchscreens and optical-scan machines. Computer-based methods are subject to “hacking,” that is, the replacement of legitimate vote-counting software with a computer

* Eugene Higgins Professor of Computer Science, Princeton University.
** Professor of Statistics, University of California, Berkeley.
program that changes (some fraction of) the votes in favor of the hacker’s preferred candidate(s). Hacking can be performed remotely (even if the machines are supposedly “never connected to the Internet”) and it is very difficult to detect. Voters and election administrators see nothing out of the ordinary.

The vulnerability of computers to hacking is well understood. Modern computer systems, including voting machines, have many layers of software, comprising millions of lines of computer code; there are thousands of bugs in that code.1 Some of those bugs are security vulnerabilities that permit attackers to modify or replace the software in the upper layers, so we can never be sure that the legitimate vote-counting software or the vote-marking user interface is actually the software running on election day.2

One might think, “our voting machines are never connected to the Internet, so hackers cannot get to them.” But all voting machines need to be programmed for each new election: They need a “ballot-definition file” with the contests and candidate names for each election, and lists of the contests different voters are eligible to vote in. This programming is typically done via removable media such as a USB thumb drive or a memory card. Vote-stealing malware can piggyback on removable media and infect voting machines—even machines with no network connection.3

There is a way to count votes by computer and still achieve trustworthy election outcomes. A trustworthy paper trail of voter selections can be used to check, or correct, the electoral outcomes of the contests in an election.

---

1 Estimates of software defect rates range from one per thousand lines of code (in high quality commercial products) down to 0.1 per thousand lines of code in extremely high-quality products (this is at the 90th percentile for the software industry). Mel Llaguno, Synopsys, Inc., 2017 Coverity Scan Report: Open-Source Software—The Road Ahead 16 (2017), https://www.synopsys.com/content/dam/synopsys/sig-assets/reports/SCAN-Report-2017.pdf [https://perma.cc/H8L3-SADH]. Modern voting machines contain software components such as an operating system (e.g., Windows 7 is fifty million lines of code; or Linux is twenty-seven million lines). USB drivers (common on voting machines) are quite large software components and are riddled with insecurities. Dave Tian et al., SoK: “Plug & Pray” Today – Understanding USB Insecurity in Versions 1 through C, IEEE Symp. Security and Privacy (2018). Therefore, we can expect 100 to 1000 bugs per million lines of code; some small portion of these are “exploitable vulnerabilities,” that is, an adversary can exploit them to take over the computer and install fraudulent software. A software-based product such as a voting machine can be expected to contain, at any given time, one or more exploitable security vulnerabilities.


“Electoral outcome” means the winning candidates or positions,\(^4\) not an exact numerical tally.

The principle of “evidence-based elections”\(^5\) is that local election officials should not only find the true winner(s) of an election, but they should also provide the electorate convincing evidence that they did. Generally, that means that eligible voters must have had the opportunity to vote, the election must have used voter-verified paper ballots, there must be convincing evidence that those ballots were kept inviolate through the audit, and the reported outcomes must be checked against the paper trail by suitable audits or hand counts.

To have affirmative evidence that reported outcomes are correct requires conducting elections using an auditable voting system, then auditing the results appropriately. First, we discuss auditability, or the creation of a trustworthy paper trail. Second, we discuss auditing—the method for efficiently assessing whether the computer-reported election outcomes are correct, based on the paper trail.

II. VOTER-VERIFIED PAPER BALLOTS

Society wants evidence that election outcomes are correct (e.g., the candidate actually selected by the voters wins the election), even if the computers have been hacked. The only known practical way to have trustworthy ballots to audit, even if the computer software has been hacked, is to have paper ballots, marked with the voters’ choices, that are manually interpretable, accountable, auditable, and re-countable.

A. Hand-Marked Paper Ballots (Optical Scan)

The traditional method of creating this paper trail (since about 1890 in the U.S.) is the use of a preprinted ballot form that lists, for each contest, the names of the candidates. Alongside each candidate is a target (square, oval, etc.) in which the voter indicates a vote. In recent decades, as such ballots are counted by optical scanners, the voter is asked to fill in an oval or complete an arrow to indicate selections. This is a hand-marked paper ballot.

With a hand-marked paper ballot, the marks on the ballot necessarily reflect what the voter did, and we can have reasonable assurance that the human-readable mark on the ballot is for the candidate actually intended by the voter. This assurance increases if the ballot follows standard best-practice ballot-design guidelines, such as those published by the U.S. Election

\(^4\) Or, for instance, whether there is a runoff.

Assistance Commission. Voters are more likely to overlook certain contests on the ballot, to overvote, to undervote, and to make other mistakes if the ballots do not follow these design guidelines.

Hand-marked paper ballots can be quite accurate: in the 2008 Minnesota election for U.S. Senator, of 2.4 million votes cast, only 0.01% (1 in 10,000) was so ambiguous that the State Canvassing Board could not interpret it, and the optical-scan voting machines agreed with the hand-recount totals with an accuracy of 99.99%.

B. Direct-Recording Electronic (DRE) Machines

Direct-recording electronic voting machines have a user interface (typically a touchscreen) and an internal computer. Voters indicate their votes on the touchscreen, and the computer program interprets those indications to add votes to counters in its memory. At the close of the polls, the computer outputs the results by printing them on paper and saving results to a removable-media cartridge.

With a DRE, the record of the vote does not necessarily reflect what the voter did. If the DRE is “hacked,” that is, if fraudulent software is installed, then the fraudulent computer program can report arbitrary fraudulent votes. There is no effective paper trail. Results may be printed on paper at the close of polls, but this paper trail starts only when the computer program is reporting the totals. That printout can be effective in auditing the aggregation of votes from different precincts, but it cannot serve as a check on the computer program in the voting machine. This is a fatal flaw of paperless DRE voting machines.

In the early 21st century, many states used DRE voting machines. But, because of the widespread recognition of this fatal flaw, only a handful of states use paperless DRE voting machines, and many of those states are transitioning to technologies that have a paper trail starting from the individual voter’s ballot.

C. Voter-Verifiable Paper Audit Trail (VVPAT)

In the 2000s, it was thought that a good solution to the problem of DREs was a voter-verifiable paper audit trail. For a DRE with VVPAT, the

---


voter indicates choices on a DRE touchscreen. Then, the DRE prints the voter’s selections on paper, behind glass. The voter inspects (“verifies”) the VVPAT; and the VVPAT serves as the ballot of record in recounts or audits.

As we discuss below, VVPAT is not an adequate solution: in practice, the vast majority of voters do not verify the paper printout—it is “voter verifiable” but not “voter verified”; and the few who do inspect the VVPAT cannot safeguard the votes of their fellow voters who do not.

D. Ballot-Marking Devices (BMDs)

Ballot-marking devices have a user interface (typically a touchscreen) on which voters indicate their selections; then the BMD prints a paper ballot that will be optically scanned. There are many variations of this technology: the paper ballot may have only human-readable marks or the votes may also be encoded in barcodes. The paper ballot might print a summary of the voter’s selections, or also contests the voter skipped. The paper ballot may be displayed under glass; it may be ejected for the voter to hold and inspect before feeding back into a slot for scanning; or it may be ejected for the voter to carry to a separate optical scanner.

None of these designs is trustworthy. Just like DRE VVPATs, a BMD vote record might not reflect what the voter did. BMDs print out a paper ballot that is, in principle, voter verifiable, but is not, in practice, voter verified. In a study of voters using BMDs in a 2018 election in Tennessee, DeMillo et al. found that 47% of voters did not inspect their BMD-printed ballots at all; the other 53% looked at their paper ballot for an average of 3.9 seconds, not nearly long enough to check that the printout matched what they indicated on the touchscreen for all 18 contests on the ballot. In a controlled experiment with real voters (but not in a real election), Bernhard et al. found that when the BMD deliberately mis-recorded one vote on each ballot, only 7% of the voters noticed.

If a BMD is hacked and systematically steals 5% of the votes in one contest and only 7% of voters inspect their ballots carefully enough to notice, then the effective rate of vote-theft is 5% × 93%, or 4.65%; this is enough to change the outcome of a moderately close election. The same analysis applies to a DRE+VVPAT system.

One might think: “not everyone needs to carefully verify their ballots;” if only 7% of voters carefully inspect their ballots, they can serve as a kind of

“random audit” of the BMDs. But this sentiment fails to hold up under careful analysis.\textsuperscript{10} If and when a voter observes that the BMD-printed ballot is marked with votes that they did not intend, the voter is supposed to alert a poll worker, who is required to void that ballot and allow the voter to mark a fresh ballot. But this situation does not provide usable evidence that the BMD was cheating: the voter might be mistaken or lying.\textsuperscript{11}

Therefore, in our hypothetical scenario in which a hacked BMD steals 5\% of the votes, and 7\% of voters carefully inspect their ballots (and know what to do when they see a mistake), then 7\% \times 5\% of voters will alert a poll worker; that is, 1 in every 285 voters will claim their paper ballot was mismarked—if the voters do not assume it was their own error. The BMD would successfully steal “only” 4.65\% of the votes.

One might think: “but some voters caught the BMD cheating, red-handed.” But nothing can be done. It is a rare election official who would invalidate an entire election because 1 out of 285 voters complained.\textsuperscript{12}

The gap between voter verifiable and voter verified makes BMDs unacceptable; hacked BMDs can steal the vast majority of the votes they set out to steal, before those votes are recorded onto the paper trail. The same analysis applies to a DRE+VVPAT system.

E. All-In-One BMDs

All-in-one BMDs combine the ballot-marking functionality of “pure” BMDs with the scanning/tabulating functionality of optical scanners. In various configurations sold by different manufacturers, the “all-in-one” or “hybrid” BMD may eject the ballot for the voter to inspect before feeding it back into the slot from which it was ejected, or the BMD may display the ballot under glass for voter inspection before retracting it past a scanner.

These machines are even less secure—and less acceptable for use in public elections—than pure BMDs. The same paper path contains both the printer (for marking ballots) and the optical scanner (for scanning ballots). The legitimate software (installed by the manufacturer) presumably will not print additional votes onto the ballot after the voter has inspected it, but hacked software could. The software installed on the BMD has complete control over


\textsuperscript{11} See Matthew Bernhard et al., Public Evidence from Secret Ballots, in 10615 ELECTRONIC VOTING, 84, 86 (Robert Krimmer et al. eds., Springer 2017); Tyler Kaczmarek et al., Dispute Resolution in Accessible Voting Systems: The Design and Use of Audiotegrity, in 7985 E-VOTING & IDENTITY 127, 131 (James Heather et al. eds., Springer 2013).

\textsuperscript{12} See generally Bernhard et al., supra note 9.
all the physical functions of the paper path: printing, scanning, and paper transport. Therefore, the hacked computer can print votes on the ballot after the voter’s last opportunity to inspect the paper. Even those 7% of voters who carefully inspect their ballots are not safe. The same analysis applies to a DRE+VVPAT system.13

F. Internet Voting

Internet voting cannot be secured by any currently known technology.14 Even if a cryptographic protocol is used to attempt to create an audit trail, the end-user device (phone, computer, or kiosk) is easily hackable. Thus, the voter may indicate a vote for one candidate, but the vote that is encrypted, authenticated, and transmitted may be for another candidate. End-to-end cryptographic paperless voting protocols are an interesting topic for future academic research, but their security and practicality is not mature enough for use in public elections. These scientific facts are well established;15 we do not discuss them further here.

G. Software Independence, Contestability, Defensibility

By 2004 it was recognized by most experts that paperless DREs were subject to a massive security hole: if fraudulent software was installed in them, that software could steal votes without any way to detect or correct the fraud, or a trustworthy way to recount. In 2008, this understanding was framed in the term “software independence”:

An undetected change or error in its software cannot cause an undetectable change or error in an election outcome.16

This notion is essential, but still too weak. It very much matters who detects the change, and the consequences of this detection. For instance, if an individual voter (amongst the approximately 7% who carefully inspect their BMD-printed ballots) detects an error, that voter has effectively detected a possible error in the election outcome. The election-outcome error is not undetectable and the system is software independent.

But in such a case, the election-outcome error is, for all practical purposes, undetectable and uncorrectable by election officials. Voters cannot

13 See also Appel et al., supra note 10.
15 Id.
prove that the votes printed on the paper are not the same as the ones they selected on the BMD. Without any such proof, it would be irresponsible to have a do-over election just on the say-so of a few individual voters.

Appel, DeMillo, and Stark propose the terms “contestable” and “defensible” as more useful in the analysis of voting-system security:

A voting system is contestable if, when an undetected change or error in its software causes a change or error in an election outcome, the system can always produce public evidence that the outcome is untrustworthy.

A voting system is defensible if, when the reported electoral outcome is correct, it is possible to generate convincing public evidence that the reported electoral outcome is correct—despite any malfunctions, software errors, or software alterations that might have occurred.

A voting system based on BMD-marked ballots is neither contestable nor defensible. A voting system based on hand-marked paper ballots, counted by optical scanners and re-countable (and auditable) by humans, are both contestable and defensible—provided careful procedures are practiced to check administrative processes, physical chain of custody of the ballots, and other physical security measures. Such procedures are called compliance audits.

III. RISK-LIMITING AUDITS

If there is a trustworthy paper record of the votes—meaning that a full, accurate hand tabulation of the recorded votes would show the true winners—there is a way to check whether the computers misbehaved: count the votes by hand.

That is an expensive prospect, so some states mandate looking at a sample of ballots instead, i.e., auditing. Generally, statutory audits provide no assurance that, if a reported outcome is wrong, the error will be detected, much less corrected.

---

17 Appel et al., supra note 10.
18 Some officials claim that the statutory audits check whether the machines are working correctly. But machines never work perfectly. The question is whether they worked well enough, in this election, to find the true winner(s). That is the question a risk-limiting audit answers.
In contrast, a “risk-limiting audit” (RLA) is any post-election procedure that offers the following statistical guarantee:\footnote{Risk-limiting audits have been endorsed by the National Academies of Sciences Engineering, and Medicine, the Presidential Commission on Election Administration, the American Statistical Association, the League of Women Voters, Common Cause, Verified Voting Foundation, and many other organizations concerned with election integrity. RLAs have been piloted more than fifty times in thirteen U.S. states and in Denmark. They are required by statute in Colorado, Nevada, Rhode Island, and Virginia, and authorized by statute in California and Washington. RLAs were developed in 2007; the first publication is P.B. Stark, \textit{Conservative Statistical Post-Election Audits}, 2 \textit{Ann. Appl. Statistics} 550 (2008). Since then, there have been extensions for other social choice functions (e.g., proportional representation), see P.B. Stark, \& V. Teague, \textit{Verifiable European Elections: Risk-limiting Audits for D'Hondt and Its Relatives}, 3 \textit{JETS: USENIX J. ELECTION TECHE. \& SYS.} 18 (2014) (for auditing any number of contests simultaneously, for different types of voting equipment, etc.); see also P.B. Stark \& M. Lindeman, \textit{A Gentle Introduction to Risk-Limiting Audits}, 10 \textit{IEEE SECURITY \& PRIVACY} 42, (2012) (for a general but still somewhat technical introduction); P.B. Stark, \textit{Sets of Half-Average Nulls Generate Risk-Limiting Audits: SHANGRLA VOTING ’20} (forthcoming 2020) [hereinafter Stark, \textit{Half-Average Nulls}] (for the most recent and efficient methods for RLAs).}

If the reported electoral outcome is wrong, there is a known, pre-determined minimum chance that the procedure will correct the reported outcome.

The maximum chance that the procedure will \textit{not} correct the outcome, if the outcome is wrong, is the “risk limit.” For instance, an RLA with a risk limit of 5\% has at least a 95\% chance of correcting the reported outcome if the reported outcome is wrong (and no chance of altering a correct reported outcome).

The only possible touchstone for determining the correct outcome and correcting wrong outcomes is the paper trail: an RLA corrects the outcome by conducting a careful, full manual tally of the paper trail. The result of that tally replaces the reported outcome if the two differ.

If the paper trail is trustworthy—i.e., if a full hand tabulation would show who really won—the replacement outcome is the correct electoral outcome, and the overall procedure limits the risk that an incorrect reported outcome will become official. If the paper trail is not trustworthy (for instance, if it has not been kept secure or if it was generated by BMDs), no procedure can limit the risk that an incorrect reported outcome will become official. Indeed, applying an RLA procedure to an untrustworthy paper trail could even replace a correct reported outcome with an incorrect outcome. At best, applying an RLA procedure to an untrustworthy paper trail can check whether tabulation error altered the outcome reflected in the untrustworthy paper trail.
There are many methods for conducting risk-limiting audits. For instance, a full hand count is a risk-limiting audit, with a risk limit of zero. But, by inspecting randomly selected ballots and using appropriate statistical methods, it is possible to conduct risk-limiting audits much more efficiently—provided the reported electoral outcome is correct.\(^\text{20}\)

IV. COMPLIANCE AUDITS

An RLA procedure that relies on an untrustworthy paper trail, or any audit that purports to ascertain voter intent from an electronic record or from an artifact that the voter did not have the opportunity to check, is “security theater.” There is little reason to believe that a full manual tally of such records would reveal the true winner(s). It is therefore crucial to base audits on voter-verified paper records; to ensure that those records include every validly cast vote exactly once, and no other votes (checking the determination of eligibility, in particular); to ensure that those records remain complete and intact from the moment they are cast through the audit; and to assess the evidence that they are trustworthy. Absent affirmative evidence that the paper trail is a trustworthy record of voter intent—i.e., that tabulating it accurately would show who won according to the intent of every voter who legitimately cast a ballot (in the contests under audit) and no others—the audit might be likely to confirm the incorrect outcome or to change a correct outcome into an incorrect outcome.

The process of assessing the trustworthiness of the paper trail is called a “compliance audit.” Compliance audits should include the following steps, among others:

**Ballot Accounting.** Check that the number of ballots sent to polling places equals the total number of cast ballots, spoiled ballots, and unvoted ballots. For systems that print ballots on demand, check that the paper stock of voting sheets cast, spoiled, and still blank adds up to the number of sheets sent to the polling place or vote center. Use accountable ballot stock, rather than plain paper, as an important security measure. Check that the number of ballots returned from each polling place does not exceed the number of voters registered at that polling place or the number of pollbook signatures at the polling place. Check that the number of ballots of each style corresponds to the number of ballots of each style reported by the voting system. Ballot counts for this purpose should be based on the physical paper, not on

---

\(^\text{20}\) When the reported outcome is incorrect, the audit is *intended* to have a large probability of requiring a full manual tally, so it generally will not save labor in that case.
the voting system: the audit needs external touchstones to check the voting system.

**Eligibility.** Check signature verification on vote-by-mail ballots. Check the disposition of provisional ballots to ensure that all that were validly cast (and no others) were included in the results. Check that each voter received the correct ballot style based on each voter’s eligibility. For vote-by-mail ballots, there should be a record of the ballot style mailed to the voter; for in-person voting, this might require recording (e.g., in pollbooks) the ballot style given to the voter. For provisionally cast ballots, this might be more complicated.

**Physical chain of custody.** Adopt a formal seal-use protocol\(^{21}\) for the tamper-evident seals on ballot boxes and other important records, e.g., use numbered, tamper-evident seals that are hard to forge or bypass; train staff in assessing evidence of tampering; record seal numbers when seals are applied; and check seal numbers against records. Review custody logs. Check that at least two staff members accompanied the ballots whenever ballots were not locked securely or under surveillance. Review surveillance video of the secure ballot storage facility to ensure there was no unauthorized access to ballots.

**Due diligence regarding processes, equipment, etc.** Review voting equipment event logs. Review any complaints made by voters or anomalies or problems noted by poll workers.

Some of these steps are formally or informally part of the canvass procedure in some jurisdictions. Ideally, the Secretary of State would require these steps (and others) to be conducted in a way that is publicly verifiable and would require jurisdictions to publish the results. Before the election, voter registration databases should be scrutinized and changelogs included. Pre-election “logic and accuracy testing” should include compliance review of the ballot design against EAC usability guidelines\(^{22}\) to ensure that voters will understand the ballot and will not inadvertently overlook some contests or mark ballots incorrectly.

---


Compliance audits should be a standard part of any recount and not just a precursor to risk-limiting audits. Absent a compliance audit, there is little reason for the public to trust that a recount will find the true winner(s).

V. EFFICIENT RISK-LIMITING AUDITS

The basic strategy behind current methods for risk-limiting audits begins by acknowledging that the reported electoral outcome might be incorrect, then examines randomly selected ballots until either (a) the evidence is convincing that a full manual tally would confirm the reported outcome, or (b) there has been a full manual tally.

There is more than one way to do this. Two basic building blocks are ballot-polling and comparison. Both can be conducted by randomly selecting either groups of ballots (batch-level audits) or individual ballots (ballot-level audits).

Ballot-polling audits are like exit polls, but instead of asking voters how they voted, the audit manually examines randomly selected ballots. If a sufficiently large sample of ballots shows a sufficiently large margin in favor of the reported winner, that is evidence that the reported winner really won. Ballot-polling audits have the advantage of requiring very little of the voting system: just the reported winners and access to the ballots. They also require local election officials to organize the ballots well enough to draw a random sample of ballots.

Comparison audits compare how the voting system tallied groups of ballots to how humans tally the same physical group of ballots. A group might be, for instance, all ballots tallied in a given precinct or by a given machine, which yields a batch-level comparison audit. The most efficient comparison audits use groups consisting of individual ballots, which yield ballot-level comparison audits. To conduct a ballot-level comparison audit, the system must report how it interpreted individual ballots in a way that allows the corresponding physical ballot to be identified and retrieved for manual inspection. Such interpretations are called “cast-vote records” or CVRs. The CVR for a ballot lists the voting system’s interpretation of voter intent for each

---

23 Ballot-level audits tend to require examining fewer ballots total than audits based on larger batches. Roughly speaking, the number of batches one needs to examine to confirm a contest with a given margin of victory at a given risk limit is about the same, regardless of the batch size. Hence, to attain a given risk limit, an audit that uses batches the size of precincts (say, 500 ballots per batch on average) requires examining about 500 times as many ballots as an audit that uses batches consisting of a single ballot (i.e., a ballot-level audit).

24 Unlike people, ballots always reply, and always reply truthfully, so ballot-polling audits give strong statistical evidence while exit polls generally suffer from large biases.

25 How to quantify the strength of the evidence depends on how the sample is drawn, among other things.
contest on the ballot. Most legacy voting systems cannot report CVRs in a way that the corresponding ballot can be identified and retrieved, but some newer systems have this capability.

One method for conducting a ballot-level comparison audit with a 5% risk limit requires manually inspecting approximately \(7/(\text{diluted margin})\) ballots, unless the audit finds errors in the CVRs. The “diluted margin” is the margin of victory in votes, divided by the total number of ballot cards\(^2\) in the population from which the sample is drawn (which must include all ballot cards cast in the contest, and may include others). For instance, in the 2018 gubernatorial primary in California, Newsom and Cox advanced to the general election. The margin of Cox over Villaraigosa, the runner-up, was 618,215 votes out of 7,060,646 ballots cast, including undervotes. The diluted margin is thus 618,215/7,060,646 = 8.76%. A ballot-level comparison audit with a risk limit of 5% would have required inspecting approximately \(7/0.0875 = 80\) ballots selected at random from the entire state (assuming the audit did not find any errors). A ballot-polling audit with a risk limit of 5% would have been expected to examine 443 ballots (assuming that the reported results are correct). For either approach, the amount of work required to justify public confidence in the outcome is *de minimis.*

Most ways of conducting RLAs require a “ballot manifest” describing how ballots are stored. For example, “There are 913 boxes of ballots, numbered 1 through 913. Box 1 contains 301 ballots. Box 2 contains 199 ballots . . . .” It is reasonable to require local election officials to construct ballot manifests routinely—if election officials cannot keep track of how much paper there is and where it is, they are not doing their job. Some counties might not currently organize their paper flow in a way that makes constructing ballot manifests possible.

\(2\) A “ballot” often consists of two or more “ballot cards” that contain different contests. Sorting the physical ballot cards into homogeneous groups can greatly reduce the number of cards that must be inspected at random to yield a given number of cards that contain a particular contest.
Ballot manifests should be constructed without relying on the voting system to count the paper; otherwise, we are trusting the voting system to check itself.27 28

VI. RESOURCES FOR RISK-LIMITING AUDITS

Ballot polling requires a ballot manifest and the reported results—the hardware and software requirements are minimal, and open-source code exists for all the computations.29 Batch-level comparison RLAs using precincts as batches generally do not save effort compared to ballot-polling RLAs for typical margins and precinct sizes, but require substantially more “data wrangling.” Ballot-level comparison audits require voting systems that can report cast-vote records for individual ballots in a way that allows the corresponding physical ballot to be retrieved, and vice versa; however, most current voting systems do not have this ability. Ballot-level comparison audits also require exporting those CVRs and “committing” to them in a publicly verifiable way.

RLA methods exist for all common social choice functions used in the U.S., including plurality, vote-for-n plurality (e.g., school boards), super-majority, and instant-runoff voting (IRV, also known as ranked-choice voting, or RCV), as well as proportional representation.30

There is a variety of open-source software to select random samples of ballots and perform risk calculations.31 The most difficult aspect of auditing is

27 However, ballot manifests can be augmented by data from the voting system to facilitate audits, provided the audit is designed to take into account the possibility that the voting system data are incorrect. For instance, there are ways to combine cast-vote records with ballot manifests to make it easier to sample ballots that contain specific contests and still ensure that the procedure is an RLA. See Stark, Half-Average Nulls, supra note 19; Michelle Blom et al., Sets of Half-Average Nulls Generate Risk-Limiting Audits (SHANGRLA), GitHub, https://github.com/pbstark/SHANGRLA [https://perma.cc/WN2U-FJGU] (last visited Mar. 17, 2020).

28 Moreover, common human errors include scanning the same box of ballots twice and failing to scan a box of ballots. Scanner mis-picks and errors resulting from clearing scanner paper jams can also cause the number of actual ballots to differ from the number according to the voting system. Relying on the voting system to construct a manifest would miss such errors. See, e.g., P.B. Stark, Tools for Ballot-Polling Risk-Limiting Election Audits, UNIV. CAL., BERKELEY: DEP’T OF STATISTICS, https://www.stat.berkeley.edu/~stark/Vote/ballotPollTools.htm [https://perma.cc/DBS2-8LXB] (last modified Feb. 16, 2017).

29 See Stark, Half-Average Nulls, supra note 19.

30 See Stark, Half-Average Nulls, supra note 19.

logistical: coordinating audits of contests that cross jurisdictional lines. That can be facilitated by well-designed software.

In our experience, it takes about two minutes to retrieve a particular randomly selected ballot and transcribe the votes for two or three contests. Additional contests take on the order of ten seconds each per audited ballot. The cost of conducting RLAs seems to be very small compared to the overall cost of holding an election. In Colorado, some local election officials report that RLAs are easier than the statutory audits that RLAs replaced, even though the previous audits had little evidentiary value.

A. Audit the Digital Images?

Some vendors are promoting systems that create digital images of ballots. These vendors claim that the images make RLAs easier to perform because fewer (or no) paper ballots need to be inspected. That is incorrect: if a risk-limiting audit relies on images of ballots, it must check that the error in making the images from the voter-verified paper ballots plus the error the system made interpreting those images to make cast-vote records is not large enough to cause the electoral outcome to be wrong. It is a mathematical fact that this requires examining at least as many physical ballots as an audit that compares CVRs to a human reading of the paper ballots, without relying on the digital images.\(^{33}\)

VII. **Principles for Election Integrity Legislation**

Laws to ensure that election results are trustworthy should satisfy a number of principles:

\(^{32}\) The process is much faster if serial numbers are printed on the ballots (after the voted ballot has been dissociated from the voter’s identity).

\(^{33}\) See supra note 28 for errors that could result in missing or duplicated images. Moreover, there are demonstrations that scanners can inadvertently alter images in ways that would change the appearance of voter intent, including erasing votes. Expecting digital images to accurately reflect voter intent from every validly cast ballot, exactly once, is wishful thinking, even in the absence of hacking. Of course, hacking the scanners or the image processing software is within the technical ability of many undergraduate computer science students.
1. **Require rigorous physical custody of ballots, and compliance audits, as discussed above.** A RLA that relies on an untrustworthy paper record accomplishes little.

2. **Require genuine RLAs.** The procedures and calculations should ensure that whenever an outcome is incorrect, the audit has the requisite chance of leading to a full hand count.\(^\text{34}\) This entails a number of things:

   a. **The audit must ascertain voter intent manually—directly from the human-readable marks on the paper ballots the voters had the opportunity to verify.** It is not adequate to rely on digital images of ballots, printout from an electronic record, barcodes, or other artifacts that are not verifiable by the voter or are not tamper evident. Nor is it adequate to re-tabulate the votes electronically, either from images of the ballots or from the original paper. BMD printouts, digital images of ballots, re-printed ballots, and other computer data are not reliable records of voter intent. They can be incomplete, fabricated, or altered (accidentally or maliciously) by software bugs, procedural lapses, or hacking. Statutes should prohibit relying on such things for the determination of voter intent. Making this prohibition explicit is important because, as mentioned above, voting system vendors are marketing technology that purports to facilitate RLAs by allowing auditors to examine digital images of ballots instead of paper ballots. Relying on an electronic record created by the voting system to accurately reflect voter intent amounts to asking a defendant whether the defendant is guilty.

   b. **The audit must take all validly cast ballots into account.** If ballots are omitted from consideration, for instance vote-by-mail ballots that did not arrive by election night or provisionally cast ballots, the audit cannot be a genuine RLA. Still, there are ways to begin an RLA before all ballots are available.

   c. **The audit must have the ability to correct incorrect outcomes.** This might mean that the audit must take place before results are certified or that the audit can revise already-certified results.

\(^\text{34}\) The statute should not dictate methods or calculations, only principles. This makes it possible to use improved methods as they are developed or as voting systems are replaced.
3. **Set the risk limit in statute.** Allowing the Secretary of State or local election official to choose the risk limits may create a real or apparent conflict of interest.

4. **Specify how the contests to be audited are selected.**
   
a. **If not every contest will be audited in every election, the selection of contests to audit should involve a random element to ensure that every contest has some chance of being selected.** This ensures that a malicious opponent will not be able to predict whether any particular race will be audited.

b. **Every contest not audited with an RLA should be audited using a risk-measuring audit instead.**

   
c. **Statutes must require RLAs on cross-jurisdictional contests—including statewide contests.** Because the point of an RLA is to ensure that reported contest outcomes are correct, every county involved in a particular contest must examine ballots in such a way that the overall cross-jurisdictional procedure is an RLA of that contest. Operationally, auditing cross-jurisdictional contests requires coordination among counties, so each county knows when its portion of the audit can stop. For example, the Secretary of State can tell each jurisdiction how many ballots it needs to draw from each cross-jurisdictional contest in light of the margin and what the audit reveals as it progresses.

5. **The audit sample must not be predictable before the audit starts.** Otherwise, any hacked software would know in which precincts it is safe to cheat. Audits in Colorado, California, Rhode Island, and elsewhere have initialized a random number generator by rolling dice in a public ceremony to ensure that the sample is unknown until that time.

---

35 Risk-measuring audits are related to risk-limiting audits, but they do not have a pre-specified minimum chance of requiring a full manual tabulation when that tabulation would show a different result. In statistical terminology, a risk-measuring audit reports a $P$-value for the hypothesis that a full count would yield a different electoral outcome, based on the audit data. Equivalently, it reports the smallest value for which a risk-limiting audit conducted using that value as its risk limit would have stopped without examining more ballots.

The sample from any collection of ballots should not be selected before election officials have “committed” to the tally of those ballots. For example, nobody should be able to know whether precinct 207 will be audited until the election official has published the tally for precinct 207.\textsuperscript{37}

6. **The public must be able to verify, not merely observe, that the RLA did not stop prematurely.** Among other things, this requires election officials to: disclose the algorithms used to select the sample, calculate the risk and determine when the audit can stop; provide the public the opportunity to observe the selection of the “seed” for drawing the sample; provide adequate public evidence that the paper trail of cast ballots is complete and intact (evidence generated in part by the compliance audit); provide the public the opportunity to verify that the correct ballots were inspected during the audit; provide the public the opportunity to observe the voters’ marks on the ballots that were inspected by the audit;\textsuperscript{38} and in “ballot-level comparison audits,” provide the public proof that the correct cast-vote record was compared to each audited ballot and proof that the full set of cast-vote records yields the reported contest results.

VIII. CONCLUSION

Electronic records of ballots are easy to manipulate by computer hacking. Therefore, voter-verified paper ballots must serve as the auditable evidence that connects the voters’ selections with the election outcome.

Optical scan voting systems, using hand-marked paper ballots designed with usability in mind, have proved to be reliable and highly accurate. These voting systems should be used with compliance-auditable ballot accounting and chain-of-custody procedures, coupled with risk-limiting

\textsuperscript{37} There are examples (notably, in Cuyahoga County, OH) where election officials altered tallies in precincts selected for recount after the sample was selected to ensure that the inspection would not find any discrepancies. See Kim Zetter, *The Mysterious Case of Ohio’s Voting Machines*, WIRED (Mar. 26, 2008, 5:51 PM), https://www.wired.com/2008/03/the-mysterious/ [https://perma.cc/2388-JC6G].

\textsuperscript{38} It is important to have published rules governing how marks on ballots are to be interpreted in audits and recounts. For instance, if a voter makes a write-in vote for a candidate who is also listed on the ballot, is that a valid vote? If a voter marks a vote for a listed candidate and also writes in that candidate’s name, is that a valid vote? If a voter marks a vote for a candidate, crosses through the mark, and marks a vote for a second candidate, is that a valid vote for the second candidate? If a voter makes a stray mark on the ballot that is distinctive enough to identify the ballot, is the ballot valid?
audits of election tallies, to achieve reliable and trustworthy evidence-based elections.

BMDs were originally envisioned as assistive devices for voters with disabilities who are unable to mark a paper ballot with a pen. Such BMDs have touchscreens, audio interfaces, and ports for other assistive technologies.

Only recently, some states and counties have adopted voting systems that use BMDs for all voters. In light of the insecurity of BMDs—the chasm between voter-verifiable and voter-verified BMD ballots—hand-marked paper ballots should be the default option presented to all voters, with BMDs available to voters who wish to use them.

Most states already use paper ballots; what we now need to conduct evidence-based elections is better procedures for safeguarding ballots, compliance audits, and risk-limiting audits. These procedures should be enacted in statutes so they have sufficient force of law to truly safeguard our elections against software hacking, insider manipulation, and other threats.

Deploying RLAs (and associated compliance audits) involves the coordination of statistical methods, administrative procedures, paper handling, and more by election administrators across towns, counties, and states. This cannot be done overnight: it requires developing methods appropriate to the election procedures in each state, training officials, educating citizens, practice, and experience. For these reasons, the National Academies of Sciences report recommends that states and local jurisdictions begin with pilot programs and work toward full implementation.39

---

39 NAT’L ACADS. SCI., ENG’G, & MED., supra note 2.
WHY INTERNET VOTING IS DANGEROUS

Barbara Simons*

CITE AS: 4 GEO. L. TECH. REV. 543 (2020)

TABLE OF CONTENTS

I. IF I CAN BANK ONLINE, WHY CAN’T I VOTE ONLINE? ...................... 543
II. HOW SERIOUS IS THE PROBLEM? .................................................. 544
III. WHAT ARE SOME OF THE RISKS OF INTERNET VOTING? ........... 546
IV. HOW SECURE ARE INTERNET VOTING SYSTEMS? ....................... 548
   A. The District of Columbia ....................................................... 548
   B. The City of Toronto ............................................................ 548
   C. Estonia ................................................................................. 549
V. WHAT ARE THE ARGUMENTS FOR INTERNET VOTING? ............. 550
   A. Military Voters .................................................................... 551
   B. Voters with Disabilities ....................................................... 552
   C. Increased Voter Participation ............................................... 554
VI. DOES BLOCKCHAIN MAKE INTERNET VOTING SAFER? ............. 555
VII. WHAT ARE THE THREATS TO BLOCKCHAIN VOTING SYSTEMS? 557
    A. Voatz .................................................................................. 558
    B. Security Issues ................................................................. 560
VIII. CONCLUSION ........................................................................ 563

Internet voting is the return of voted ballots over the Internet, using a computer, a tablet, or a smart phone. The voted ballot may be transmitted via a web portal, as a PDF or other attachment, or as a fax.

I. IF I CAN BANK ONLINE, WHY CAN’T I VOTE ONLINE?

That is a question we have heard repeatedly for many years. What questioners often do not appreciate is that we cannot bank online either—at least not with a guarantee of security. Banks will continue to provide online banking as long as it is less expensive to cover financial losses from

* Retired, IBM Research; Member, Board of Advisors of the U.S. Election Assistance Commission; Board Chair, Verified Voting. Thanks to the Georgetown Law Technology Review for their helpful edits. Any mistakes herein are my own.
cyberattacks than to build new buildings and hire new staff. The situation is nicely summarized in a report from the Atlantic Council:

When a hacker steals money online, the theft is easily discovered. Banks, online retailers, and other companies offering services over the Internet factor in some degree of loss as a cost of doing business online, and generally indemnify their customers against bad actors. Online voting poses a much tougher problem: lost votes are unacceptable. Online voting systems are complex, and any updates often must be separately recertified by election authorities. And unlike paper ballots, electronic votes cannot be “rolled back” or easily recounted. The twin goals of anonymity and verifiability within an online voting system are largely incompatible with current technologies. Russian state-sanctioned hackers, it should be recalled, brought almost all of Estonia’s online activities to a halt in 2007 and might do so for online elections as well. Nobody knows whether the DRE [voting] machines or other proprietary voting systems in use elsewhere have already been hacked too.¹

An electronic vote cannot be “rolled back” because there is no way to know if it is accurate: What the voter sees on her device’s screen may differ from what is stored in the device’s memory, sent over the Internet, or received at the polling place. Thus, the secret ballot makes it impossible for the voter to verify her ballot.²

II. How Serious Is the Problem?

For years, we have been hearing about successful attacks on a variety of institutions, such as Capital One, Google, Facebook, the FBI, Symantec, Marriott, and the Office of Personnel Management (OPM). All of these institutions have significant financial and personnel resources available to

² There is ongoing research that uses encryption to develop “verifiable end-to-end” internet voting systems. See, e.g., U.S. VOTE FOUND., THE FUTURE OF VOTING: END-TO-END VERIFIABLE INTERNET VOTING—SPECIFICATION AND FEASIBILITY STUDY (2015), https://www.usvotefoundation.org/E2E-VIV [https://perma.cc/QML9-EER3]. However, no currently available system, including blockchain voting systems, provide this type of verification. See discussion infra Part VI.
them to prevent hackers. In contrast, election officials tend to be underfunded and under-resourced with little to no cybersecurity expertise at their disposal. It is highly unlikely that most local election officials would be able to withstand attacks from powerful nation-states, or even from clever local hackers.

Intelligence experts have been warning for several years that our electoral system is under attack:

- “[T]here were multiple, systematic efforts to interfere in our election.”
  — Special Counsel Robert Mueller III

- “He [Putin] tried again to muck around in our elections this last month. We are seeing a continued effort around those lines.”
  — James Mattis, former Secretary of Defense

- “. . . Russia attempted to interfere with the last election and continues to engage in malign influence operations to this day.”
  — Christopher A. Wray, F.B.I. Director

The findings of the intelligence community were reflected in a bipartisan report issued by the Senate Intelligence Committee on October 8, 2019, which stated:

. . . DHS [Department of Homeland Security] assessed that the [Russian] searches, done alphabetically, probably included all 50 states, and consisted of research on general election-related web pages, voterID information, election system software, and election service companies.

Russia is not the only country capable of conducting cyberattacks on our elections. North Korea famously hacked Sony because of *The Interview*, a

---

comedy about an assassination attempt on Kim Jong-un.\textsuperscript{7} A Chinese state-sponsored hacking group is suspected in a cyber campaign targeting U.S. utility companies.\textsuperscript{8} On January 3, 2020, Christopher Krebs, Director of the DHS Cybersecurity and Infrastructure Security Agency, warned: "Bottom line: time to brush up on Iranian TTPs [Tactics, Techniques, and Procedures] and pay close attention to your critical systems . . . Make sure you’re also watching third party accesses!"\textsuperscript{9}

While there is no evidence that votes were changed in 2016, no proper investigation was conducted because (1) there is no national post-election ballot audit or recount, (2) most of our state laws have not been updated to reflect the risks introduced by the use of computers in elections, (3) some state laws actually appear to be designed to inhibit or prevent post-election ballot audits or recounts, and (4) it is currently impossible to recount any type of paperless voting systems, because we do not know if the results stored in the computers’ memories accurately reflect the voters’ intentions.\textsuperscript{10}

\section*{III. What Are Some of the Risks of Internet Voting?}

As an old cartoon says, “On the Internet, nobody knows you’re a dog.”\textsuperscript{11} As that cartoon suggests, one of the threats of Internet voting, or indeed of any kind of remote voting, is that the ballot could be cast by someone other than the voter.\textsuperscript{12} While illegal, this act is unlikely to be prosecuted. Likewise, since the United States does not have a national ID, it is essentially impossible


\footnotesize{\textsuperscript{10}In addition to voter marked paper ballots and laws that require post-election ballot audits and/or recounts, voters must check their paper ballots—especially if the ballots have been produced by machines—and there must be a secure chain of custody of those ballots. These are all important issues, but they are not the focal points of this paper.}

\footnotesize{\textsuperscript{11}The cartoon, by Peter Steiner, was published in \textit{The New Yorker} on July 5, 1993. \textit{See also} Glenn Fleishman, \textit{Cartoon Captures Spirit of the Internet}, N.Y. TIMES, Dec. 14, 2000, at G8.}

\footnotesize{\textsuperscript{12}Internet voting is the return of voted ballots over the Internet, using a computer, a tablet, or a smart phone. The voted ballot may be transmitted via a web portal, as a PDF or other attachment, or as a fax.}
to authenticate a potential voter. Other threats include: voter coercion and vote buying/selling, malware on the voter’s device that can change or discard the voter’s selections without the voter’s knowledge, “man-in-the-middle” attacks that intercept the voter’s ballot as it is traversing the Internet, and denial of service attacks that can prevent ballots from reaching election officials.

Despite the threats and multiple warnings, there are no regulations governing Internet voting, primarily because no one knows how to write them. In particular, there are no standards (federal or state) regarding independent testing, government oversight, legal accountability, or indeed the ability to conduct a recount.

The National Institute of Standards and Technology (NIST) was asked to develop Internet voting standards, but has not done so. Instead, NIST produced reports that warned about threats to Internet voting. One such report states:

In addition, the platforms not under the control of election officials [i.e. the voter’s computer or smart phone] may be poorly protected and vulnerable to malware, phishing, and denial of service attacks. These platforms may be the target of attacks to monitor and/or modify voter choices, capture personal information, or prevent a voter from accessing the voting services . . . . When voting platforms contain malware, the voting platform may try to inhibit a voter from casting his or her ballot, alter a voter’s choices, monitor how a voter votes, use the voter’s credential to gain and expand access to damage the voting system, change election results, or harm the credibility of the election results.

NIST has consistently warned of the dangers of Internet voting, as a recent website posting exemplifies: “Malware on voters' personal computers

13 There are claims that smart phones can be used to authenticate voters. We discuss those claims later in the paper. See discussion infra Part VI.
poses a serious threat that could compromise the secrecy or integrity of voters’ ballots.”

IV. HOW SECURE ARE INTERNET VOTING SYSTEMS?

A. The District of Columbia

The first example of an Internet voting system that was subjected to independent testing was the 2010 “digital vote by mail” pilot project in Washington, DC. The system, which was developed by the Open Source Digital Voting Foundation, aimed to provide Internet voting for UOCAVA voters.

The District of Columbia Board of Elections and Ethics (BOEE) took the enlightened and unusual path of providing a “public review period” during which anyone was allowed to attempt to break into a mock election. The public testing was scheduled to run from late September to early October. However, by October 1, voters in the mock election were hearing the University of Michigan Fight Song played after they cast their ballots. The song turned out to be the “calling card” of the university team that had successfully broken into the system within 36 hours of the start of the test. Of course, an intruder wishing to manipulate an election would not be so obvious.

The Michigan team, under the leadership of Professor Alex Halderman, was able to modify previously cast ballots, rig subsequently cast ballots, and reveal voters’ selections, thereby violating the voters’ right to a secret ballot. They even controlled the network infrastructure for the pilot and were able to watch network operators configure and test the equipment. When the Michigan team observed attempted break-ins that appeared to be from China and Iran, they protected the system from those break-ins.

B. The City of Toronto

In 2014 Toronto issued a Request for Proposal (RFP) for Internet voting which mandated that vendors competing for the contract first submit

---


17 Attempting to break into a real election, even if the intent is to expose vulnerabilities, is illegal. The information about this exercise is taken from Scott Wolchok et al., Attacking the Washington, D.C. Internet Voting System, in PROC. 16TH CONFERENCE ON FINANCIAL CRYPTOGRAPHY & DATA SECURITY 114 (2012). See also J. Alex Halderman, Hacking the D.C. Internet Voting Pilot, FREEDOM TO TINKER (Oct. 5, 2010), https://freedom-to-tinker.com/2010/10/05/hacking-dc-internet-voting-pilot/ [https://perma.cc/HA2U-ECN5] for a less technical explanation.

18 Wolchok et al., supra note 17, at 10.
their systems for security examination by independent experts. Since vendors typically do not provide their systems for independent inspection, the Toronto RFP—which should be typical—was quite unusual.

The independent experts recommended against the purchase of any of the submitted systems:

Of the proposals evaluated in the context of the RFP process, it is our opinion that no proposal provides adequate protection against the risks inherent in internet voting. It is our recommendation, therefore, that the City not proceed with internet voting in the upcoming municipal election.\(^{19}\)

The Toronto study is one of the few examples of an independent assessment of commercial Internet voting systems.

C. Estonia

In March 2007 Estonia became the first country to authorize Internet voting in a national parliamentary election. (Voters still had the option of casting paper ballots.) In April 2007, Estonia suffered massive cyberattacks that in some cases lasted weeks and appeared to have originated in Russia.\(^{20}\)

While Estonia is often used as an example of how Internet voting can work, this depiction fails on several counts, as discussed below.

Estonian cryptographer Helger Lipmaa explained why he was casting a paper ballot:

Voter computers are an obvious problem: most of the people are computer illiterate, and are not able to check if their computers are not infected. Even if they have the newest antivirus (which we can’t be sure of), that antivirus itself might not be able to detect a piece of new malware that has been written specifically for *that* election and is unleashed just before it.\(^{21}\)


In 2011, the mayor of Tallinn, who also was the country’s first prime minister, invited me to visit Estonia. As the leader of Estonia’s second largest political party, the Centre Party, he and his colleague were concerned that Internet voting was being used to undermine election results in favor of the Reform Party. While I could not determine whether or not election rigging had occurred, during my visit, I expressed a number of concerns relating to the vulnerability of voters’ computers to election rigging malware, insider threats, the possibility that the system could have been attacked by anyone anywhere, the lack of transparency (the software was not publicly available), and the lack of a security evaluation by independent computer security experts.\(^{22}\)

More precise threats were uncovered in 2014 when a group of independent security experts studied the system and concluded that:

\[
\ldots \text{there are multiple ways that state-level attackers, sophisticated online criminals, or dishonest insiders could successfully attack the Estonian I-voting system. Such an attacker could plausibly change votes, disrupt elections, or cast doubt on the integrity of results. These problems are difficult to mitigate, because they stem from basic architectural choices and fundamental limitations on the security and transparency that can be provided by procedural controls. For these reasons, we recommend that Estonia discontinue the I-voting system.}\(^{23}\)
\]

Perhaps most disturbing is the distrust by a major Estonian political party of election results engendered by Internet voting. Because it is impossible to validate the results of an Internet election, it also is impossible to determine whether or not election rigging has occurred. This is a very unhealthy situation for any democracy, especially a relatively new one.

V. WHAT ARE THE ARGUMENTS FOR INTERNET VOTING?

Proponents of Internet voting argue that it is needed for overseas military voters, voters with disabilities, and to increase voter participation.


A. Military Voters

In 1986 Congress passed the Uniformed and Overseas Citizens Absentee Voting Act (UOCAVA). Consequently, military and overseas voters often are referred to as UOCAVA voters.

The 2009 Military and Overseas Voter Empowerment Act (MOVE) significantly sped up ballot delivery to UOCAVA voters. MOVE requires states to make blank ballots available electronically at least 45 days before an election. The voter can download the ballot, print it out, mark it, and return the voted ballot by postal mail. MOVE also provides free expedited mail service for voted ballots of overseas uniformed service voters. While there are security risks with the online posting of blank ballots, those risks are dwarfed by the risks of returning voted ballots over the Internet.

MOVE has made it possible for almost all military voters to return their voted ballots in a timely fashion, as was confirmed by an analysis conducted shortly after the passage of MOVE by the Military Postal Service Agency (MPSA):

> Election officials must adhere to the MOVE Act requirement for dispatching absentee ballots to voters no later than 45 days prior to the election date. This provides adequate time for ballots to reach absentee voters in the most remote locations to vote and mail back their ballot for the election.

Inevitably, there will be a small number of service people who will be unable to return their voted ballots in a timely fashion. These cases are sometimes used to demand Internet voting for UOCAVA voters—instead of calling for extending the date for receipt of voted ballots, as has been done by several states. As a result, and despite multiple warnings, Internet voting is allowed in approximately 30 states, primarily for UOCAVA voters.

---

B. Voters with Disabilities

The 2002 Help America Vote Act (HAVA), which allocated almost $4 billion for the purchase of new voting systems, also required that polling places provide an accessible voting system for voters with disabilities.\(^{27}\) In part because of the explicit mention of Direct Recording Electronic (DRE) machines in HAVA, the bulk of the early post-HAVA systems were DREs that stored the voted ballot in the memory of the machine. Many of these systems were paperless, providing no opportunity to check the accuracy of the results.\(^{28}\)

Vendors claimed that the DREs were secure, though we have known since the first independent security study of DREs that they have significant vulnerabilities.\(^{29}\) Vendors also claimed that the DREs were easy for voters with disabilities to use, thereby satisfying the HAVA requirement for accessible voting systems.\(^{30}\) The accessibility claim, however, turned out to have been an exaggeration at best. For example, the California Secretary of State’s 2007 Top-to-Bottom Review of California voting machines examined the accessibility of those machines:

> Although each of the tested voting systems included some accessibility accommodations, none met the accessibility requirements of current law and none performed satisfactorily in test voting by persons with a range of disabilities and alternate language needs.\(^{31}\)

---


\(^{28}\) When there was an outcry against the paperless DREs, vendors developed a retrofit called Voter Verified Paper Ballots (VVPATs). For a host of reasons, including that voters tended not to check the VVPATs, these machines turned out to be, at best, marginally better than the paperless DREs. We still are coping with the HAVA legacy today. Only Louisiana remains entirely paperless as of this writing, but many other states have paperless jurisdictions, making it impossible to conduct post-election ballot audits or recounts in those states. See The Verifier—Polling Place Equipment—November 2020, VERIFIED VOTING, https://www.verifiedvoting.org/verifier/ [https://perma.cc/K5SN-BS8M].


\(^{30}\) 52 U.S.C. § 21081 (a)(3)(A) (“The voting system shall be accessible for individuals with disabilities, including nonvisual accessibility for the blind and visually impaired, in a manner that provides the same opportunity for access and participation (including privacy and independence) as for other voters . . . ”).

\(^{31}\) NOEL RUNYAN & JIM TOBIAS, ACCESSIBILITY REVIEW REPORT FOR CALIFORNIA TOP-TO-BOTTOM VOTING SYSTEMS REVIEW 1 (2007),
Vendors producing new Ballot Marking Devices (BMDs) also state that the BMDs have good accessibility features. While there has not been much accessibility testing of commercially available BMDs, a Pennsylvania Department of State examination of the ExpressVote XL manufactured by ES&S, the largest national voting system vendor, found three classes of accessibility problems and concluded that “[v]erification is possible, but challenging.” These included display problems that do not make the text sufficiently large for low vision users, a lack of information for blind voters such as announcing the party of each candidate, and an inability of blind voters or those with severe vision impairment to verify their ballots. For example, because the XL “displays the printed ballot under a glass panel, and then casts the ballot by automatically depositing the paper ballot in a container while it records the vote electronically,” it is “impossible for voters to use personal technology such as magnifiers or text readers to read the paper ballot.”

Not surprisingly, a number of disability rights advocates, as well as Internet voting proponents, argue that Internet voting will facilitate voting by those with disabilities. They observe that voters with disabilities can be confronted with numerous obstacles if they attempt to cast their ballots at the polls. These include polling places that are difficult (or even impossible for wheelchairs) to navigate, a lack of facilities for voters with mobility limitations, and voting systems that fail to provide adequate disability accommodations.

Fortunately, we do not need to repeat the mistakes of the past by providing insecure or inadequate voting systems for voters with disabilities, namely Internet voting. Technology exists that allows voters with disabilities to download a blank ballot, mark it from home using their accessibility technology, print the voted ballot, and then return it via postal mail. While


33 Id. at 19.


35 See blind voter Noel Runyan’s description of voting on an early DRE in id. at 216–17.

36 For voters with disabilities who choose to vote in person at a polling place, there is at least
not perfect, this is far more secure than Internet voting and is deployed in several states, including Oregon and California.\footnote{See Our Story, \textsc{Five Cedars Group}, http://www.fivecedarsgroup.com/#ourstory [https://perma.cc/EL4N-YYLK].}

C. Increased Voter Participation

The argument for Internet voting that appears to have the most general appeal is that it will increase voter participation. However, not only is there no evidence that Internet voting increases voter participation, there is substantial evidence to the contrary.

The parliament of British Columbia, Canada, allocated roughly $420,000 (Canadian dollars) for a study on Internet voting.\footnote{Indep. Panel on Internet Voting B.C., Recommendations Report to the Legislative Assembly of British Columbia 7 (2014), https://elections.bc.ca/docs/recommendations-report.pdf [https://perma.cc/2VLV-VGWU].} The report recommended against Internet voting at the time it was written and warned that “There are significant risks to implementing Internet voting that can jeopardize the integrity of an election . . .”\footnote{See id. at 47.} Among other findings, it dispelled the myth that Internet voting increases voter participation:

While there have been some Internet voting elections where voter turnout has increased, when other factors such as the apparent closeness of the race and interest in particular contests (e.g., a mayoral election without an incumbent) are taken into consideration, research suggests that Internet voting does not generally cause non-voters to vote. Instead, Internet voting is mostly used as a tool of convenience for individuals who have already decided to vote.\footnote{Id. at 12.}

Much to many people’s surprise, the report did not find that Internet voting increased participation by young people:

Researchers have also looked at the demographics of Canadian voters who have used Internet voting and have found that Internet voting is most popular among middle-age voters and

---

one BMD that prints a full size paper ballot with the voter’s selections that is designed to look like a hand marked paper ballot, thereby making it difficult to distinguish ballots marked by voters with disabilities from other ballots and protecting the secret ballot. See Marketing Brochure, Verity Touch Writer: Ballot Marking Device by Hart InterCivic (2016), https://www.hartintercivic.com/wp-content/uploads/VerityTouchWriter.pdf [https://perma.cc/FQ7E-7BUD].
least popular among youth and therefore reflects traditional voter turnout demographics. These findings run contrary to the widely expressed belief that Internet voting will lead to increased participation by youth.41

Another notable example is that of Estonia, which has allowed Internet voting since 2005. Recent figures show that voter turnout has declined from 60.6% in 2009 to 53.3% in 2017; turnout in the last Parliamentary election also declined from 64.2% in 2015 to 63.7% in 2019.42

In another example, no significant impact on turnout was detected when Switzerland allowed Internet voting in the cantons of Geneva and Zurich in its federal elections from 2004 to 2014.43

In sum, Internet voting appears to be a solution in search of a problem—it is not needed for military voters, there are better and safer options for voters with disabilities, and it does not appreciably increase voter participation.

VI. DOES BLOCKCHAIN MAKE INTERNET VOTING SAFER?

A blockchain is a distributed data structure that could have single or multiple owners. In the case of multiple owners, a majority must agree before a transaction is added to the blockchain. Both ownership arrangements have their own vulnerabilities. For example, colluding owners could determine which transactions are added, including possibly false ones. Furthermore, outside attackers who penetrate the servers containing the blockchain might be able to manipulate transactions. But because, blockchain voting is likely to have a single owner—namely the local election official or the vendor—any advantage of having multiple owners that keep a check on each other is eliminated.44

41 Id. at 13 (emphasis added).
But one of the largest issues with blockchain voting is that, at its core, blockchain voting is still Internet voting. The National Academy of Sciences produced a report on voting security that warned about these issues:

Conducting secure and credible Internet elections will require substantial scientific advances. The use of blockchains in an election scenario would do little to address the major security requirements of voting, such as voter verifiability. The security contributions offered by blockchains are better obtained by other means. In the particular case of Internet voting, blockchain methods do not redress the security issues associated with Internet voting.

As suggested by the National Academy, since a voted ballot is transmitted over the Internet, blockchain voting is Internet voting, regardless of what vendors claim. Nonetheless, one of the largest blockchain voting vendor, Voatz, refers to blockchain Internet voting as “mobile” (as opposed to “traditional”) voting and downplays the Internet voting aspect:

While there are different definitions that may come to mind for “Internet Voting”, the term typically refers to a browser residing primarily on a voter’s PC connected over the Internet to a web server. There are several key differences between traditional Internet voting and Voatz. First, only recently-manufactured smartphone models from Apple, Samsung and Google are supported with Voatz. These devices are built with security features, like fingerprint and facial recognition, that extend far beyond standard browsers running on a potentially-compromised PC for voter authentication. Second, modern smartphones provide hardware-based security to store private keys which, in turn, allow highly secure, encrypted transactions to be conducted over the public Internet. Third, votes are stored on a permissioned blockchain that will eventually be controlled by various stakeholders (e.g. a Secretary of State or a state


45 See JEFFERSON, THE MYTH OF “SECURE” BLOCKCHAIN VOTING, supra note 44.
board of elections) to ensure their tamper resistance and immmutability. 47

The fact that voters are casting their ballots using smartphones does not address the multiple security threats of Internet voting. While the smartphones mentioned by Voatz use biometrics to authenticate the users, neither the authentication nor the encryption protect against potential vote rigging malware on the phone, as appears to have happened with Jeff Bezos. Bezos claims that malware was introduced into his Apple iPhone X by Mohammed bin Salman. 48

Furthermore, the fact that a cell phone recognizes its user’s biometrics does nothing to ensure that the voter is who she claims to be while authentication questions in general raise privacy concerns. Because we do not have the infrastructure to securely authenticate a remote voter, and because personally identifiable information is needed in order to validate a voter, it is likely that authentication and privacy will be linked in ways that put the voter’s private information at risk, especially with smart phone voting. Passwords are not reliable, since most users’ passwords are insecure and relatively easy to break. Much personal information, such as SSN, driver’s license number, birthdate, etc. has been stolen in massive data breaches and can be purchased, making this information untrustworthy as a tool to validate voters online. The biometrics used by a smart phone to validate the phone owner is not an option, because the biometric information is stored in the phone’s memory to which election officials do not have access—nor should they for privacy reasons.

VII. WHAT ARE THE THREATS TO BLOCKCHAIN VOTING SYSTEMS?

Voatz had dominated the market of blockchain voting on smart phones, though that may change because of some recent negative security reports. 49 I discuss how Voatz purports to work, those security reports and additional warning signs, and where the industry is moving now, below.

A. Voatz

A voter using Voatz casts her vote on her smart phone.\(^{50}\) Her voted ballot is sent over the Internet to one or more of the thirty-two Voatz blockchains, half of which are on the Microsoft Azure service and half on Amazon Web Services. There is no publicly available information as to how well those blockchains are protected against attack.

Voatz has not stated publicly what the contents of the blockchains are. For example, in addition to the actual ballots, Voatz has not stated what information about the voters, if any, it stores together with the ballots. Voatz claims to use cryptography, but they do not name the encryption scheme used nor where it is used in the process.

Once the election is over, the ballots in the blockchain are decrypted and sent to the local jurisdictions, where they are then printed out and included with the other ballots. Voatz provides no explanation of how this process works.

Voatz asserts that the voter can verify her ballot, but again it does not explain how that is done, nor even what is meant by “verify.” It appears that the voter is supposed to be able to determine what votes were recorded for her in the blockchain, but this is not at all clear. Nor is there any description of how the verification is conducted. Is the voter’s identity or some other form of ID address stored with her ballot? If so, how is the secrecy of her ballot protected? If she attempts to verify her ballot, how is that information transmitted to her?

Voatz claims to produce a paper trail:

\[\text{A paper ballot is generated on election night for every mobile vote recorded on the blockchain and the printed ballots are tallied using the standard counting process at each participating county. This also facilitates a post-election audit by comparing the paper ballots with the anonymized voter-verified digital receipts generated at the time of vote submission.}\] \(^{51}\)

The paper trail/audit claim is grossly misleading since, as we have observed, there is no way to know that the paper printout of the mobile vote accurately reflects the voter’s selections.


\(^{50}\) Much of what follows was found by parsing the Frequently Asked Questions page on Voatz.com and through my own observations. Frequently Asked Questions, supra note 47.

\(^{51}\) Id.
Voatz has never been certified, not least of all because there are no certification standards for Internet voting. However, Voatz does collect and transmit voted ballots, and therefore is a critical voting system component.

Voatz may have a significant negative impact on voters’ privacy. Voatz claims to have a method of authentication that involves sending a live facial video together with a photo of the voter’s passport photo page or driver’s license (front and back) to Jumio, a Palo Alto, California company.\(^5^2\) Jumio also collects a vast amount of personal information from the voter including name, address, birthday, driver’s license or passport number, smart phone number, a copy of the voter’s signature, etc.\(^5^3\) The data collection raises a host of questions, including: Who has access to the data? How secure is it? How long is it retained? What rights does the voter have to determine how the data is used? Is it shared with any other entity? What legal responsibilities do Voatz and Jumio have if the data is inappropriately used, sold, or stolen?

Jumio deploys machine learning to authenticate the voter, but we don’t have data about the accuracy of the authentication.\(^5^4\) Does it have different success rates on different groups, such as people of color? What recourse is there if the voter is wrongly rejected (false negative)? How likely is the software to wrongly authenticating a voter (false positive)? Is the data encrypted? If so, what type of encryption is used? Apparently, there can be human intervention in the case of a failure to match, but details of what that entails are not provided.

Perhaps most disturbing is the Jumio User Information License. Here is part of that license:

Customer hereby grants to Jumio a license to use, reproduce, modify, create derivative works from, distribute, perform, transmit, anonymize and display the User Information (including any rights specifically pertaining to biometric information) necessary to develop, provide and improve the Services, including the right for Jumio to grant equivalent rights to its service providers that perform services that form part of or are otherwise used to perform the Services.\(^5^5\)

\(^{52}\) SPECTER, KOPPEL, & WEITZNER, supra note 49, at 5.

\(^{53}\) Id. at 6.

\(^{54}\) See Jumio Corporation, About, JUMIO.COM, https://www.jumio.com/about/ [https://perma.cc/8GU9-XJRF].

Even though each voter is probably unaware of Jumio’s role in the process, Jumio’s license purports to give the company control over that voter’s data.\textsuperscript{56} Though Jumio claims compliance, it is unclear if these license conditions comply with California’s new privacy law in practice.\textsuperscript{57}

The Jumio license also forbids its customer—Voatz—from reverse engineering of the system for any reason, which would include (unauthorized) independent security testing. Further, Jumio’s license also obligates Voatz to prevent other persons—which including voters—from accessing Jumio’s product, even if accessing is part of an individual’s effort to delete that individual’s personal information.\textsuperscript{58}

Despite the lack of transparency, Voatz has conducted “pilots” (with real ballots in real governmental elections) in West Virginia during the 2018 primary and midterm and the City and County of Denver municipal general election in 2019. In both cases participation in the pilots was available to UOCAVA voters only. We do not know whether any of the pilot elections was audited—and if so, what the results were.

B. Security Issues

In early 2020 the MIT News Office announced that some MIT researchers had discovered vulnerabilities in Voatz software that would allow “hackers to alter, stop, or expose how an individual user has voted.”\textsuperscript{59} Voatz responded by calling the researchers’ report “flawed.”\textsuperscript{60} Based on the MIT study, West Virginia terminated their relationship with Voatz and instead announced that they would be using Democracy Live for their online voting.\textsuperscript{61}

Voatz also took the unusual step of hiring an outside security firm, Trail of Bits, to assess their product. Still more uncommon, Voatz released the report when it was finalized, even though it was very negative:

\textsuperscript{56} See \textsc{Specter, Koppel, & Weitzner}, supra note 49, at 7.
\textsuperscript{58} See \textit{Jumio Terms and Conditions v5.1, supra note} 55, at §§ 3.4, 3.6.
\textsuperscript{60} \textit{Voatz Response to Researchers’ Flawed Report}, \textsc{BLOG @ VOATZ} (Feb. 13, 2020), https://blog.voatz.com/?p=1209 [https://perma.cc/FZ2J-8BHR].
Our security review resulted in seventy-nine (79) findings. A third of the findings are high severity, another third medium severity, and the remainder a combination of low, undetermined, and informational severity.\textsuperscript{62}

Voatz, which is funded by venture capital firms including one associated with Overstock.com, has received substantial financial help from Tusk Holdings and Tusk Philanthropies, founded by its entrepreneur Bradley Tusk.\textsuperscript{63} The President of Tusk Philanthropies is Sheila Nix, former Chief of Staff to Dr. Jill Biden. Nix continues to push for expanding blockchain and mobile voting.\textsuperscript{64} Tusk Philanthropies was instrumental in helping to arrange Voatz’s pilot in West Virginia both by advocating for the company and by providing $150,000 for the test.\textsuperscript{65} But West Virginia is not the only “pilot”; Tusk Philanthropies has stated that as of May 2020 they have “successfully completed fourteen pilots in five different states.”\textsuperscript{66}

It appears that Tusk Philanthropies may have turned away from Voatz, because of the bad security assessments, and instead is supporting another Internet voting system called OmniBallot, produced by Democracy Live.\textsuperscript{67} In February, 2020 Tusk Philanthropies partnered with the King County (Seattle) Elections to support an online voting election using Omniballot in a local board supervisor election.\textsuperscript{68}

However, Omniballot has some of the same issues as Voatz. An Omniballot FAQ makes the false claim that “OmniBallot is not an online voting system” because “a paper ballot is downloaded by the elections


\textsuperscript{65} Halpern, supra note 63.


Therefore, a “voter verified paper ballot is always available for a hand recount if necessary.” Like Voatz, the “voter verified” claim is misleading, since it is impossible for the voter to verify the downloaded ballot that she never even seen.

Democracy Live has not provided security-related details to independent cybersecurity experts. They claim that an “independent audit” by the nonprofit National Cybersecurity Center (NCC) found that the King County election was accurately tabulated and that there was no interference. But no report was publicly released. Moreover, while the CEOs of both Voat and Democracy Live are on the NCC’s Secure the Vote Advisory Board, there is a notable absence of election security experts on the Security Board. Vendor involvement raises questions about the independence of the NCC.

Despite Democracy Live’s secrecy, in June 2020 a security analysis of Democracy Live’s online voting system was released. The analysis showed that the system was “vulnerable to vote manipulation by malware on the voter’s device and by insiders or other attackers who can compromise Democracy Live, Amazon, Google, or Cloudflare. In addition, Democracy Live, which appears to have no privacy policy, receives sensitive personally identifiable information—including the voter’s identity, ballot selections, and browser fingerprint—that could be used to target political ads or disinformation campaigns.”

Because of the lack of openness on the part of Democracy Live, the researchers were able to analyze only the application software used by the

---


73 Id. at 1. A week after Spector and Halderman published their report, Democracy Live (finally) posted a privacy policy. See Privacy Policy, DEMOCRACY LIVE (last updated June 15, 2020, 8:00 A.M.), https://democracylive.com/privacy-policy/ [https://perma.cc/X8WP-LNXZ]. But Democracy Live still collects the voter’s name, physical address, email address, telephone number, partial social security number, and information about the phone or computer used to access the site.
voters, and not the software that runs the backend servers.\textsuperscript{74} Not long after the security report was released, Delaware initially decided not to use the Democracy Live system for their July 7 primary.\textsuperscript{75} Delaware subsequently backtracked by allowing the use of the Democracy Live system coupled with the return of voted ballots via mail, fax, or email, disallowing only casting a voted ballot via the website. The Delaware Department of Elections wrongly claimed that “no votes are cast online under any circumstances”\textsuperscript{76} We note that email voting, which has risks similar to website voting, involves sending voted ballots over the Internet, and faxes typically are sent unencrypted, often over the Internet.

\textbf{VIII. Conclusion}

The warnings of attacks on our elections have not ceased, with the Department of Homeland Security sounding the alarm: “Russia, China, Iran, and other foreign malicious actors all will seek to interfere in the voting process or influence voter perceptions.”\textsuperscript{77} More than ever, we cannot afford to put our democracy at risk by indulging in Internet voting, especially since there are far safer options.

The burden of proof to show that any Internet voting system is safe should be the responsibility of whoever is advocating for such a system, including policy makers, citizens, and vendors. Until we are provided with such proof, Internet voting—including blockchain voting—should not be deployed in any governmental election.


\textsuperscript{76} See SPECTER \& HALDERMAN, supra note 72, at 21 (Figure 4).

NEW CHALLENGES IN ELECTION OVERSIGHT & REGULATION

Our fourth and final panel of the day considered regulatory solutions to the problems that were identified throughout the symposium. The panel critically interrogated various approaches to regulating technology-based election interference, including transparency-based approaches such as the Honest Ads Act, privacy-based approaches such as the Voter Privacy Act of 2019, proposed federal and state bills regulating bot speech, and emergency planning standards that could address the worst-case scenarios.

Alexandra Givens, then Director of Georgetown’s Institute for Technology Law and Policy, moderated this panel. She was joined on stage by Marc Lawrence-Apelbaum of the Campaign Legal Center, Patrick Day formerly of the Senate Judiciary Committee and currently with Cloudflare, the Honorable Karen Kornbluh of the German Marshall Fund of the United States, and the Honorable Ellen Weintraub of the Federal Elections Commission.

In the following pages, we are proud to include articles by Mr. Lawrence-Apelbaum, Mr. Day, Ambassador Kornbluh, and Commissioner Weintraub. And even though she did not directly participate in this particular panel, we are also pleased to include Professor Julie Cohen’s article on platform regulation in this section.
DESPERATE TIMES CALL FOR DESPERATE MEASURES: THE NEED FOR FEDERAL LEGISLATION TO ADDRESS NEW POTENTIAL ELECTION DAY THREATS

Marc Lawrence-Apfelbaum*

CITE AS: 4 GEO. L. TECH. REV. 567 (2020)

TABLE OF CONTENTS

I. BACKGROUND .................................................................................................................. 568
II. THE CURRENT LEGAL FRAMEWORK ......................................................................... 572
III. PROBLEMS WITH THE CURRENT LEGAL FRAMEWORK ........................................ 574
IV. POTENTIAL FIXES ........................................................................................................ 576
   C. Use a New York Statute as a Model to Provide a Framework for Post-Election Day Responses ........................................................................................................ 576
   D. Use a Double-Trigger Mechanism to Postpone Elections Before or On Election Day ...................................................................................................................... 577
   E. Statutory Time Frames Should be Modified to Allow for More Time for Rescheduled Elections ........................................................................................................ 579
V. CONCLUSION .................................................................................................................. 581

* Marc Lawrence-Apfelbaum recently served as a special advisor on foreign election interference and online disinformation threats to U.S. elections at the Campaign Legal Center (CLC), where he completed a project on practical ways that CLC could expand its work to confront cyberthreats to democracy. This article is derived from some of that work. Marc thanks Richard Clarke and Emilian Papadopoulos of Good Harbor Security Risk Management; Norman Ornstein of the American Enterprise Institute and Board Chair at CLC; Meredith McGehee of Issue One; Professor Alex Halderman of the University of Michigan; Alan Frumin, former Parliamentarian of the U.S. Senate; and John MacGaffin, his mentor on this mission. Their insights and advice were invaluable to him in his project and writing this article. He also thanks the editors of the Georgetown Law Technology Review. Before CLC, Marc was the long-tenured EVP, General Counsel & Corporate Secretary of Time Warner Cable Inc., where he founded and chaired the company’s Cybersecurity Council and dealt with cutting-edge online legal and policy issues. J.D. Georgetown Law; B.A. University of Pennsylvania.
I. BACKGROUND

More than thirty election-security bills have been introduced in Congress, but most have been stalled in the Senate and no major legislation has been enacted.1 The bills would require states to make greater use of paper ballots and paper trails, adopt effective audit procedures, and strengthen federal cybersecurity support for state and local election authorities. While almost all these bills would be useful, it is probably now too late for states to make significant changes in the election equipment they use for the 2020 elections.2

Federal, state and local authorities have also increased contingency planning to deal with, and prepare for, potential election day crises.3 Among the possible crisis scenarios that have been “war-gamed” are the following:

- Hacks of voting equipment that could change actual vote tallies;
- Hacks of voter registration rolls that could disrupt voting or falsely indicate that particular voters have already cast their ballots;4

---

2 Although Congress approved an additional $425 million late in 2019 for additional funding for states to upgrade voting equipment, it came “many days late and many dollars short according to experts, who say billions were needed starting at least two years ago.” Editorial Board, There’s a lot to like in Congress’s new election security measures. But there’s a big omission., WASH. POST (Dec. 25, 2019, 6:22 PM), https://www.washingtonpost.com/opinions/the-spending-bills-glaring-omissions-on-election-security/2019/12/25/24495180-25b4-11ea-ad73-2fd294520e97_story.html [https://perma.cc/D5ZJ-3NP2]. In addition, “states are generally reluctant to undertake major election equipment replacements within a year of major elections or primaries.” See, Lawrence Norden, How to Secure Elections for 2020 and Beyond, BRENNAN CTR. FOR JUST. (Oct. 23, 2019), https://www.brennancenter.org/our-work/research-reports/how-secure-elections-2020-and-beyond [https://perma.cc/4UYS-4WWZ].
4 Although many states that employ electronic poll books at polling places to check whether voters are registered now keep paper backups to use if that equipment becomes unavailable, it may not be possible in such circumstances to tell which voters have already cast their votes. While a wholesale switch to provisional ballots might be theoretically possible in such circumstances, the long lines that would result would likely lead many voters to go home without voting. See THE BLUE RIBBON COMMISSION ON PENNSYLVANIA’S ELECTION SECURITY, STUDY AND RECOMMENDATIONS 50
• Attacks on the electrical grid to shut down polling sites and election systems;
• Social media disinformation campaigns that would confuse voters about the places and methods for voting and create fake news about attacks or disasters on election day;\(^5\)
• Hacks of state and local election websites that provide information on polling place locations and hours of operation;
• The insertion of ransomware onto election equipment and systems; and
• Many other potential disruptive scenarios, along with the recognition that it is impossible to predict or mitigate them all ahead of time.\(^6\)

Some election experts think that the possibility of foreign state-actor intrusions into election equipment that would actually change election outcomes is fairly remote.\(^7\) Among other considerations, they believe that precautions have been buttressed since 2016, changing actual votes is hard, and doing enough of it to matter would be harder still.\(^8\)

---


\(^6\) In 2014, long before most Americans had ever heard of the Internet Research Agency, which led some of Russia’s major social media disinformation campaigns in the 2016 presidential election, the IRA apparently engineered a sophisticated fake news operation about an invented chemical explosion in St. Mary Parish, Louisiana. The tactics included the manufacture of CNN web pages and video reports, and it led to widespread confusion. See Adrian Chen, *The Agency*, N.Y. TIMES MAG. (June 2, 2015), https://www.nytimes.com/2015/06/07/magazine/the-agency.html [https://perma.cc/3NEU-VGGV]. Although many wondered at the time why the Russians would have been interested in running such an operation in a small town in Louisiana, it now seems evident what the reason was: practice.

\(^7\) For a discussion of possible election-day crises, see, e.g., Joseph Marks, *The Cybersecurity 202: Feds and police are war-gaming all the ways an election can be hacked*, WASH. POST (Nov. 5, 2019), https://www.washingtonpost.com/news/powerpost/paloma/the-cybersecurity-202/2019/11/05/the-cybersecurity-202-feds-and-police-are-war-gaming-all-the-ways-an-election-can-be-hacked/5dc0a27b88e0fa10f6d20b60/ [https://perma.cc/RG7U-3TFV].

As discussed in numerous government reports and press accounts, however, the goal of the Russians, and probably other state actors, is not to tamper with actual votes. Rather, the goal is to generally sow confusion and doubt in the U.S. and other democracies about the legitimacy of the democratic system itself, including taking steps that undermine the confidence in and outcomes of our elections.9 Using the tactics in the cyber-scenarios set out above on election day would be far easier to accomplish and could inflict as much or greater harm as would changing vote counts.10 Moreover, the use of these techniques in a limited number of key districts would likely be as impactful as broader attacks. The large-scale planning for these events suggests that the possibility of them occurring is far greater than the possibility of actual vote changes.11

The list of foreign adversaries that likely will employ these tactics has also grown since 2016. It now includes China, North Korea, and Iran.12 Indeed, following the U.S. assassination of an Iranian general, Iran should be prioritized as an election interference threat.13

In addition to the newer risks set out in the scenarios above, there has always been the risk, and actual occurrence, of hurricanes and other natural disasters taking place on or around election days. There have also been many examples of “hanging chads” and other incidents that have led to election


10 See, e.g., Marks, supra note 6 (“[H]ackers could destroy public faith in an election’s outcome without changing any votes. And that’s particularly concerning because many of these targets are far more vulnerable than voting machines.”)

11 See, e.g., Marks, supra note 6.


recounts. The 9/11 terrorist attacks occurred on a primary election day in New York and led to a state-ordered do-over of the elections held that day. In such circumstances, one might have thought that federal and state election laws would be clear on how to deal with natural disasters and terrorist attacks that occur on election days and might also provide guidance on how to proceed in the face of the newer threats described above. An examination of existing federal and state law, however, indicates that federal election day emergency law does not exist, that some, but not all, states have laws that purport to address emergency matters, and that the states that have addressed this issue take a variety of approaches.

In particular, existing state laws vary greatly in terms of which government officials have the authority to postpone elections or order recounts, the allowable bases for doing so, and the timeframes in which this must occur. To further complicate things, some of those timeframes could result in postponed elections or recounts occurring beyond the dates by which federal office holders’ terms expire. The timelines could also result in delays that could prevent states from taking advantage of the federal safe harbor that makes state determinations of its electors for president incontrovertible by Congress if they are certified by a certain date.

While some experts worry that too much public discussion of these issues might itself feed into the goal of foreign adversaries to delegitimize election results and democracy, doing nothing to confront these problems could result in far worse outcomes. Though the traditional approach of leaving elections largely in the hands of the states might make sense in dealing with

---

14 This article focusses on postponements not recounts, but as with postponements, no federal legislation governs recounts; not all states have statutes that do; there are many different approaches among the state laws that do exist; and there are similar potential timing issues as set out below regarding postponements. See generally Automatic Recounts, NAT’L CONF. ST. LEGISLATURES (Jan. 30, 2020), http://www.ncsl.org/research/elections-and-campaigns/automatic-recount-thresholds.aspx [https://perma.cc/CV2B-TRLQ].


16 See infra note 25.

17 The “safe harbor” statute provides that, if there is any dispute over which electors’ votes should be counted, Congress must accept as conclusive any final determination under state law that is made at least six days before the date fixed before the meeting of electors. This means that it must be made within thirty-five days of election day. 3 U.S.C. § 5 (2018); see also Daniel P. Tokaji, An Unsafe Harbor: Recounts, Contests, and the Electoral College, 106 MICH. L. REV. FIRST IMPRESSIONS 84 (2008), https://repository.law.umich.edu/mlr_fi/vol106/iss1/14/ [https://perma.cc/K7KB-VG66]. The Supreme Court’s decision in Bush v. Gore, 531 U.S. 98 (2000) was issued on the safe-harbor date, and “was partially predicated on Florida’s intent to avail itself of [it].” Id.
natural disasters where risks and responses may vary based on local conditions, the newer election day crises are likely to be similar wherever they occur. Inconsistent local responses, including none at all, could greatly impact public confidence in elections. Even if there are no statutory guidelines for these crises, legal challenges would still be filed in a fashion similar to Bush v. Gore, but on a potentially far greater scale and leave even more confusion in its wake.

II. THE CURRENT LEGAL FRAMEWORK

There are many intersecting provisions in the Constitution and federal and state laws that address the workings of elections for federal offices, including their timing, postponements and recounts.

The Constitution provides that each state may prescribe its own rules for the “Times, Places, and Manner” of elections to the House and Senate. But the very next clause takes this back, providing that “Congress may at any time by Law make or alter such Regulations, except as to the Places of chusing (sic) Senators.” Although the wording that sets out Congress’s authority over state elections for presidential electors is somewhat different, the Constitution also gives Congress full authority to set the date for such elections. Taken together, these provisions give Congress full “residual and superseding authority” over most election issues, including their timing and postponement.

Despite Congress’s broad authority to override state laws concerning elections for federal offices, Congress has generally left such matters to the states. The most notable exceptions for this article are: Congress has set dates

---


19 That Congress cannot override the place for “chusing” of Senators is probably a remnant of the original framework in the Constitution, before the 17th Amendment provided for the popular election of Senators, that they were elected by their state legislatures. The Constitution also provides that each house retains the authority to be the final judge of the elections of its members. U.S. CONST. art. I, § 5 cl. 1.

20 U.S. CONST. art. II, § 1, cl. 4. (“Congress may determine the Time of chusing the Electors, and the Day on which they give their votes; which Day shall be the same throughout the United States.”) The Constitution gives the states the authority to determine how do select electors, including by appointment, but all have chosen to use elections.


22 Although it does not address election-day crises, Congress did enact sweeping voting
for when elections to federal office must be held, established a safe harbor date by which states must certify their electors for President or risk having Congress decide which among competing slates constitute the proper electors, and, in accordance with the constitutionally prescribed number of years each office is held, set the dates and times on which the terms for president, members of Congress, and senators must begin and end.\textsuperscript{23} There are apparently no federal statutes that provide for election postponements.\textsuperscript{24}

As noted above, many states have no laws that govern postponements of elections to either federal or state offices. The state laws that have been enacted vary greatly from state to state, as is also the case with legislation governing recounts. These differences include the permitted bases for ordering postponements, the officials who are authorized to make the call, and the length of permitted postponements.\textsuperscript{25} In addition to state statutes that explicitly provide for election postponements in the event of emergencies, a number of states grant governors or other state officials the power to declare and suspend

\textsuperscript{23} Federal law provides that elections for president and Congress are set at the first Tuesday after November 1st in the applicable years. 3 U.S.C. § 1 (2018); 2 U.S.C. § 7 (2018). The term of office for the president expires on January 20th of the year it is up and on January 3rd for members of the House and Senate. U.S. CONST. amend. XX, §1. The safe harbor date for states to certify electors is within 35 days of election day. After that Congress, through a vote of each house, could determine which electors to seat.

\textsuperscript{24} See, e.g., JACK MASKELL, supra note 21. Section 302 of HAVA does provide that “[a]ny individual who votes in an election for Federal office as a result of a Federal or State court order or any other order extending the time established for closing the polls by a State law . . . may only vote in that election by casting a provisional ballot”. Help America Vote Act § 302.

\textsuperscript{25} See, e.g., L. PAIGE WHITAKER, CONG. RESEARCH SERV., RS21942, STATE ELECTION LAWS: OVERVIEW OF STATUTES REGARDING EMERGENCY ELECTION POSTPONEMENT WITHIN THE STATE (2004), https://fas.org/sgp/crs/RS21942.pdf [https://perma.cc/HF68-HZS2]. For example, Florida allows the governor to suspend elections upon declaring a state of emergency for up to ten days; in Georgia the governor or secretary of state upon a declaration by the governor of a state of emergency can postpone elections in affected areas for up to forty-five days; in New York, if the State Board of Elections determines that fewer than 25% of registered voters voted as direct consequence of a disaster, an additional day of voting must be held within twenty days; and in Maryland, upon declaring a state of emergency, the governor can postpone elections in all or part of the state until a specific new date, and state and local election officials can petition a court to take any action it considers necessary to protect the integrity of the electoral process. See NAT’L ASS’N OF SEC’YS OF STATE, STATE LAWS & PRACTICES FOR THE EMERGENCY MANAGEMENT OF ELECTIONS 15–20, https://www.nass.org/sites/default/files/Election%20Cybersecurity/report-NASS-emergency-preparedness-elections-apr2017.pdf [https://perma.cc/MF22-3YEA].
the operation of state laws in emergencies. These statutes may also designate the responsible officials with the authority to postpone elections for federal offices when emergencies are declared. It is worth noting that federal law allows the president and others to declare states of emergency as well, but none of the specified emergency powers explicitly relate to postponements of elections.

III. PROBLEMS WITH THE CURRENT LEGAL FRAMEWORK

The lack of federal law dealing with postponements and recounts in elections, and the patchwork quilt of state regulations, is not ideal as applied to traditional election day emergencies, including natural disasters and terrorist attacks. At least as to those events, however, their actual or probable

27 Id.
28 See generally BRENNAN CTR. JUST., A GUIDE TO EMERGENCY POWERS AND THEIR USE, (2019), https://www.brennancenter.org/our-work/research-reports/guide-emergency-powers-and-their-use [https://perma.cc/BQ2H-D33D]. Of the 176 emergency powers set out in the report that may become available to the president upon a declaration of emergency, only one explicitly deals with elections, but not postponement. This is in an Executive Order, issued on September 12, 2018, in which the president made a finding that there is a threat to national security based on foreign threats to election infrastructure and disinformation campaigns. After each election, it requires reports regarding foreign interference and gives the president power to impose sanctions in response to such occurrences. Exec. Order No. 13848, 83 Fed. Reg. 46843 (2018). https://www.whitehouse.gov/presidential-actions/executive-order-imposing-certain-sanctions-event-foreign-interference-united-states-election/ [https://perma.cc/BCE9-UDNR].
29 The coronavirus pandemic has put in stark relief just how inadequate state laws are in dealing with natural disasters and demonstrated the pressing need for federal legislation to address both natural and cyber threats to elections. In Ohio, because there were no clear standards for postponing the primary election that was scheduled for March 17, the governor engineered a lawsuit by private plaintiffs to seek its postponement. After the court rejected a request for a TRO, the governor delayed the primary by declaring a public health emergency. See Nick Corasaniti & Stephanie Saul, Ohio’s Governor Postpones Primary as Health Emergency Is Declared Over Virus, N.Y. TIMES, (Mar. 16, 2020) https://www.nytimes.com/2020/03/16/us/politics/virus-primary-2020-ohio.html [https://perma.cc/K3P3-JJJD]. Though they faced similar coronavirus challenges, Florida, Illinois, and Arizona went ahead with primary elections on that same day, although voting levels in many areas in those states were well below normal and eight percent of poll workers failed to show up in Miami-Dade County. Nick Norasaniti et. Al. Illinois Stumbles as States See Light Voter Turnout, With Many Ballots in the Mail, N.Y. TIMES (Mar 17, 2020) https://www.nytimes.com/2020/03/17/us/politics/primary-voter-turnout-2020.html [https://perma.cc/MV27-FQBT]. In response to the confusion, and to fears that voting in the November presidential election could be significantly impacted, Senators Klobuchar and
impact on the ability of citizens to vote, including in which precincts, is fairly straightforward. When it comes to the cyberthreats set out above, the judgment calls that must be made would be far more nuanced.

For example, if a foreign adversary were to launch a social media disinformation campaign that falsely led people to believe that there had been a terrorist attack, or that misdirected people as to how and where they should vote, the impact on voting could be broad yet not easily measured. The difficulties could be the same with attacks on the power grid, the launch of ransomware attacks, and attacks on e-polling books. And whether actual vote tallies were impacted might never be clear or at least take time to uncover. If similar election-day crises occurred in multiple states, and some postponed their elections, but others had no authority or declined to do so, the impact on public perceptions of election integrity could be acute.

As to problems with time frames, although past postponements and recounts created issues about whether the elections could be completed in time to meet federal deadlines, the time needed to wait out storms or to recover from actual attacks is also relatively straightforward. The time frames needed to deal with these newer threats are far less clear.

On the difficult problem of placing the power to order postponements, although the states take different approaches, it is often the governor who has the power to make such calls. Entrusting this authority to elected officials, especially ones who might themselves be candidates for reelection, always presents the risk of abuse. But the ability of courts and the public to evaluate whether natural disasters have occurred, and their scope, is again more straightforward than assessing the impact on voting of the social media disinformation campaigns and the other newer threats set out above. Indeed, one election-day crisis scenario that many experts fear is that our sitting president, were he to lose in 2020, would (with support from Russian disinformation efforts) claim that the loss was the result of voter fraud.30

Wyden introduced legislation that would permit the use of mail-in ballots by all voters in the country, and the recently enacted coronavirus stimulus bill included $400 million to support mail-in ballot efforts. See Maggie Miller, Senate Includes $400M For Mail-In Voting In Coronavirus Spending Deal, HILL (Mar. 25, 2020) https://thehill.com/policy/cybersecurity/489435-senate-includes-400-million-for-mail-in-voting-in-coronavirus-spending. [https://perma.cc/F3DB-8S4X].

30 These concerns have also been raised in the press and by at least one senior Democratic leader. See, e.g., Glenn Thrush, Pelosi Warns Democrats: Stay in the Center or Trump May Contest Election Results, N.Y. TIMES (May 4, 2019), https://www.nytimes.com/2019/05/04/us/politics/nancy-pelosi.html?smid=tw-nytimes&kntyp=eur[https://perma.cc/7K7G-RPAL]; Clark Mindock, What would happen if Trump refused to leave office after 2020 election loss?, INDEPENDENT (June. 16, 2019), https://www.independent.co.uk/news/world/americas/us-politics/trump-impeachment-
such circumstances, the question of where to place such power is a challenging one.

As also noted above, Congress has left most decisions about postponements and recounts in elections for federal offices to the states since the beginning of our republic. Although Congress has always had the authority to make more of these decisions itself, this long tradition of deferring to the states may make it politically untenable for it to now attempt to assert a greater level of control.31

But these cyber election-day crises will occur whether or not new laws and rules are enacted to deal with them. That there are real challenges to coming up with workable solutions is no reason to ignore them. As noted above, were such events to occur, elected officials and courts would take action to deal with them regardless of whether existing law provides clear answers. In the absence of clear statutory guidance, the chaos that would ensue could be colossal.

IV. POTENTIAL FIXES

Because dealing with cyberthreats is both complex and politically sensitive, the purpose of this Article is not to propose complete fixes, but to draw attention to them and encourage immediate attention by Congress, public-interest groups, and other parties. That said, what follows are some general ideas on possible frameworks.

A. Use a New York Statute as a Model to Provide a Framework for Post-Election Day Responses

The biggest problem to be confronted—apart from changes to actual vote tallies—is scenarios that result in significantly fewer voters casting their ballots on election day. This could be the outcome of many of the cyberthreats discussed above, including hacks that cause electrical outages in key districts, or social media disinformation campaigns that mislead voters about how and where to vote or falsely indicate that there has been a terrorist attack. It could also be due to hacks of e-polling books that lead to long lines or falsely show that many have already voted.

In such circumstances, there would be enough votes for states to certify election results. For states to certify elections based on greatly reduced voting


31 The power of states under existing law to order postponements of federal elections is also somewhat unclear. See Automatic Recounts, supra note 14. Federal legislation would make this issue moot.
levels that result from cyberattacks, however, would lead to the very delegitimization of elections and democracy that Russia and other state actors are trying to foment. Unfortunately, existing state laws that allow contests and recounts after election days could not remedy this problem, because they generally apply only when election results are very close. Therefore, they would not apply where vote counts are reduced, but the results are not close.

A New York election statute, however, does contain a good framework for providing for new elections in the event that some of the scenarios described above result in significant numbers of voters not casting their ballots. The statute requires that an additional day of voting be provided when a state or local board of elections determines that, in particular voting districts:

as the direct consequence of a fire, earthquake, tornado, explosion, power failure, act of sabotage, enemy attack or other disaster, less than twenty-five per centum of the registered voters of any city, town or village, or if the city of New York, or any county therein, actually voted in any general election. 32

The additional day of voting would occur only in districts where the 25% finding was made and would count votes that had already been cast in addition to new votes. 33

Whether 25% is the right figure is open to debate, but the approach that the statute employs would provide a good objective test for election do-overs following cyberattacks and would help to insulate such decisions from claims of political motivations.

B. Use a Double-Trigger Mechanism to Postpone Elections Before or On Election Day

Developing objective criteria to trigger the postponement of elections before or on election days in response to cyberthreats would be difficult, if not impossible. Just as with hurricanes and terrorist attacks, however, a severe cyberthreat may warrant postponing an election. For example, if false reports of terrorist attacks led many voters and poll workers to stay away from polling places, or electrical outages shut them down, going ahead with balloting might not be advisable or even possible.


33 When New York postponed its primary election on September 11, 2001, in response to the terrorist attacks on the World Trade Center, the state required a complete do-over of the election, disregarding votes that had already been cast, and conducting a new election statewide. See Nagourney, supra note 15. The approach of Section 308, enacted subsequent to 9/11, seems more appropriate.
Accurately judging in advance or on election day the likely impact of potential cyberattacks on voter turnout would necessarily require predictions and subjective judgments. Compounding this problem, as discussed above, many state statutes that provide for postponements place the authority to order postponements in governors or other elected officials who themselves may be running for office in the very elections they have the authority to delay. The subjectivity of these judgments and the potential conflicts of interest would cut against the legitimacy of elections that postponements are meant to restore.

One way to deal with these concerns would be to employ a double-trigger mechanism that would require designated officials at both the federal and state levels to authorize any election postponement due to cyberattacks. In addition to providing an added check against too easily requiring additional election days, this would also help to address concerns that new federal legislation in this area would inappropriately federalize decisions that should be left to the states.34

Even though partisanship and divisiveness are present in many states, the state-level officials selected should probably be the governor or state board of elections. It is not clear where else to place the authority at the state level, and the second trigger at the federal level should provide a sufficient safeguard against actual and perceived abuses. However, due to the extreme rancor at the federal level, it would probably be best not to designate the U.S. president as the federal official to make these calls. One approach might be to place the federal-level authority in the hands of the congressional “Gang of Eight,” which generally gets advance briefings of sensitive intelligence activities, or to designate a group from the national intelligence community itself to have this responsibility.35 To reduce conflict-of-interest concerns, there should also be a prohibition against officeholders who are up for reelection being involved in making a decision to postpone, with backup officials named to take the place of candidates in such circumstances.36

34 Although some offer more unprincipled reasons why the Republican-controlled Senate failed to pass any major election-security bills in 2019, others see the reason as the traditional Republican opposition to federalizing election law, which has traditionally been left to the states. See, e.g., Alex Pareene, The Simple, Odious Reason Mitch McConnell Opposes Election Integrity, NEW REPUBLIC (July. 31, 2019), https://newrepublic.com/article/154566/simple-odious-reason-mitch-mcconnell-opposes-election-integrity [https://perma.cc/9TUW-9NHT].
36 State-level officials running for reelection, in addition to federal ones, would also be
The postponements should be limited to areas affected by the particular threats, and makeup dates should be set to take account of federal constitutional and statutory deadlines. As discussed in the next section, however, the time frames should be adjusted to provide more time for rescheduled elections.

C. Statutory Time Frames Should be Modified to Allow for More Time for Rescheduled Elections

New federal legislation should also adjust the timing of requirements related to voting by the Electoral College. Although the Constitution requires that electors cast their vote on the same day nationwide, it leaves it to Congress to set that date. Also, although the Constitution sets January 20th of applicable years as the end date for presidential terms, it similarly leaves it to Congress to set the other relevant dates. Thus, Congress would have the flexibility to determine new timelines that would allow more time for rescheduled elections to be held.

Under the current statutory provisions, except where noted, the applicable dates for the 2020 presidential election are as follows:

- November 3 – Election Day (the first Tuesday after the first Monday in November)
- December 8 – (Six days before the meeting of the electors) – Safe harbor date for states to certify their slate of electors so that the votes of those electors are binding on Congress.
- December 14 – Electors cast their votes for president and vice president.
- January 6 – Joint session of Congress to count the votes (set by the Constitution, but subject to change by Congress)
- January 20 – Inauguration Day (set by the Constitution)

ineligible to play a role in these decisions. Even though, by the terms of the statute, the postponement would apply only to elections for federal offices, it’s likely that states would follow suit with their own elections. Mandated disqualification is necessary at both the federal and state levels to mitigate conflict-of-interest concerns.

37 These changes should apply to dates for all federal elections, including midterms. In turn, states would also likely change the dates for their own elections to avoid duplicative costs.

38 The existing tight time frames are a problem for states that provide for postponements and recounts in federal elections. Regardless of the passage of federal legislation, providing for more time would be worthwhile. As noted above, the Court’s decision in Bush v. Gore was based in part on ensuring that Florida could certify its electors by the “safe harbor” date, thus preventing Congress from being able to contest its slate of electors.
One bold change would be to move up election day by a month or even more.\(^{39}\) There is no magic to the first Tuesday after the first Monday in November. In fact, many commentators already believe that election season lasts too long, which is especially apparent when the U.S. election timeline is compared to those of other democracies.\(^{40}\)

Moving up Election Day but keeping Inauguration Day as is would lengthen the “lame duck” session of Congress following the election. Although some worry that officeholders who no longer face reelection may act inconsistently with the interests of their constituents, evidence on this is mixed and lame duck status may instead free legislators from the undue influence of special interests.\(^{41}\)

In addition to or instead of changing Election Day, the safe harbor date and the nationwide date for electors to cast their ballots could be pushed back. In particular, the value of the safe harbor provision that makes states’ slates of electors binding on Congress is doubtful. For one thing, there have only been two challenges in Congress to such slates, and both times Congress has affirmed the states’ choices.\(^{42}\) For another, the Court in \textit{Bush v. Gore} cited meeting the safe harbor date as one of the reasons for cutting short the Florida recount in the 2000 election, but citizens of that state and across the country would arguably have been better served by more accurately counting votes that determined the election’s outcome than by certifying questionable tallies.\(^{43}\) With these changes, the joint session of Congress could be kept in early January, to allow time for Congress to conduct any proceedings needed

\(^{39}\) To the extent that rescheduling Election Day is too extreme, there are other opportunities to provide additional time for postponements, such as changing the deadline for electors to cast their votes.

\(^{40}\) Ironically, the 20th Amendment, which moved Inauguration Day from March to January, was passed in part because technological innovations reduced the time needed for counting votes. \textit{See, e.g.}, Thomas V. DiBacco, \textit{The 20th Amendment’s Rush to Inauguration}, WALL ST. J. (Nov. 20, 2016, 5:19 PM), https://www.wsj.com/articles/the-20th-amendments-rush-to-inauguration-1479680395 [https://perma.cc/E8SF-FQVJ]. In modern day, different technological developments—the rise of hacking and disinformation spread via social media—tend to necessitate a longer time period between Election Day and Inauguration Day. Because the Constitution fixes Inauguration Day but not Election Day, legislation can move only the latter.


in the event that no candidate for president or vice president obtains an electoral vote majority.\textsuperscript{44}

V. CONCLUSION

In an era of widespread hacking and social media interference by foreign adversaries, there is a pressing need for federal legislation to provide a nationwide framework to respond to significant cyberattacks on or around election days that are meant to disrupt contests for federal office. Leaving such matters to the states, which have taken inconsistent approaches, and sometimes none at all, to older election day threats allows Russia and other state actors to succeed in their primary mission of sowing distrust in the American democratic system.\textsuperscript{45}

These issues are complex and politically sensitive. Thus, the intent of this article is not to offer complete solutions. Rather, its purpose is to identify the need for federal legislation capable of addressing these election day threats, to outline some preliminary suggestions, and to spur discussion and action in Congress and elsewhere about the shape of potential federal legislation.

\textsuperscript{44} The 12th Amendment to the Constitution provides that if no candidate obtains a majority of electoral votes, then the House chooses the president, with one vote per state, from the candidates with the three highest number of electoral votes. If no candidate for vice president obtains a majority, the Senate, by majority vote, chooses from the candidates with the two highest number of votes.

\textsuperscript{45} Providing for postponements in case of cyber threats wouldn’t directly address the problem of what to do if similar threats occur on the day of the new election. Since malevolent cyber activity is ever-ongoing, the possibility of repeat occurrences is likely greater than in the case of other threats, such as natural disasters. Nonetheless, since election postponements should be limited to the narrowest geographical areas possible, resources would only be expended in a small number of places, minimizing the destructive impact of additional disruptions.
CAMBRIDGE ANALYTICA AND VOTER PRIVACY

Patrick Day*

CITE AS: 4 GEO. L. TECH. REV. 583 (2020)

TABLE OF CONTENTS

I. INTRODUCTION .............................................................. 583
II. THE VOTER PRIVACY ACT .................................................. 587
    A. Findings and Sense of Congress ...................................... 588
    B. Rights Provided Under the Act ....................................... 589
    C. Definitions Included in the Act ...................................... 590
III. LEGAL CHALLENGES .......................................................... 591
    A. Limits on Quantity of Speech ........................................... 591
    B. Limits on Candidates’ Access to Data .............................. 592
IV. IS VOTER PRIVACY A COMPELLING INTEREST IN THE DIGITAL AGE? . 593
    A. Factual Basis for Voter Privacy ....................................... 594
       1. Modern Data Collection and Campaigns .......................... 594
       2. New Developments in Psychographics ............................. 596
       3. National Security .................................................... 600
    B. Rights Balancing and the Legal Argument .......................... 603
       1. The Voter, the Candidate, and the People ...................... 603
       2. Preserving a Candidate’s Access to Data ....................... 605
V. CONCLUSION ........................................................................ 606

I. INTRODUCTION

Finding a silver lining associated with Cambridge Analytica can be difficult. The notion of a private organization using licit and illicit means to

* Former Counsel for national security, U.S. Senate Judiciary Committee, J.D., Washington College of Law at American University; M.A., National Defense University. I am deeply thankful to Professor Julie Cohen and Alexandra Givens, from the Institute for Technology Law & Policy at Georgetown Law School; as well as Kelly McCluer, Jordan Patrick Cohen, Molly Rosen, Andrew Do, Florence Noorinejad, Gabriel Khoury, Haris Vrahliotis, Nicole Fulk, May Yang, and Joshua Banker of the Georgetown Law Technology Review for their helpful edits. Any mistakes herein are my own.
undermine elections around the world for profit is hard to reconcile with liberal democratic values. A brief sampling of the myriad allegations against the U.K.-based data analytics firm include: sharing detailed information on U.S. voters with Russian intelligence; working with a far-right U.K. politician suspected of Russian ties on the Brexit campaign; participating in a scheme to hack a Nigerian presidential candidate’s personal emails on behalf of “oil billionaires”; meeting with Julian Assange of Wikileaks while he possessed Hilary Clinton’s emails stolen by Russian intelligence; and secretly recording a candidate receiving a fake bribe offer in an election in St. Kitts and Nevis. Most of these activities occurred during the same period that Cambridge reportedly worked in more than forty-five elections in the United States.


6 Frances S. Sellers, Cruz Campaign Paid $750,000 to ‘Psychographic Profiling’ Company, WASH. POST (Oct. 19, 2015), https://www.washingtonpost.com/politics/cruz-campaign-paid-
including on behalf of three major-party candidates for president.\(^7\) Many of Cambridge Analytica’s former clients now occupy elected or appointed positions in the U.S. government,\(^8\) and several former employees work in American politics, including for the current President’s reelection campaign.\(^9\)

However, without the famous undercover video of Cambridge’s CEO offering to use Ukrainian prostitutes and bribes to undermine another foreign election,\(^10\) the public would not have had access to the wealth of information now available about Cambridge Analytica, particularly its attempt at “psychological persuasion”\(^11\) of U.S. voters.\(^12\) Publicly, Cambridge Analytica described its voter targeting process as follows: “[building] on top of demographic polling and traditional microtargeting with an extra dataset on personality, and we use that in order to understand the behavioral drivers that allow us to change voting behavior.”\(^13\) Privately, it used psychographic profiles on 230 million Americans to conduct what the firm called “‘psychological operations,’ or psyops – [which change] people’s minds not through persuasion but through ‘informational dominance’, a set of techniques...

---

8 The following U.S. government officials were reportedly clients of Cambridge Analytica: Ben Carson, Steve Bannon, John Bolton, Thom Tillis, and Tom Cotton. *Id.*
that includes rumor, disinformation and fake news.”

Prior to the undercover footage, and whistleblowers like Christopher Wiley, Cambridge Analytica was celebrated as an “up and coming company with technology not to be missed.”

Fortunately, Cambridge Analytica has now been subject to government investigations on five continents, documentaries, and two years of award-winning investigative journalism. Public outrage regarding the firm’s activities led many Americans to question for the first time the economic and national security costs associated with U.S. social media platforms’ data harvesting. And, after years of resisting lawmakers’ requests, Facebook CEO Mark Zuckerberg was compelled to testify before Congress for the first time.

The Cambridge Analytica affair also resulted in the introduction of the first federal legislation in the United States that would regulate the use of voters’ personal information in elections—the Voter Privacy Act of 2019. According to the bill’s sponsor, U.S. Senator Dianne Feinstein (D-CA), the Voter Privacy Act of 2019 is a direct response to Cambridge Analytica and is intended to mitigate “sophisticated online surveillance” by candidates and campaigns intended to target and influence voters’ “unique psychological characteristics.”


15 Id.

16 Cambridge Analytica and the Future of Data Privacy: Hearing before S. Comm. on the Judiciary, 115th Cong. (2018) (prepared statement of Sen. Diane Feinstein, Ranking Member, S. Comm. on the Judiciary) (“numerous governments have launched formal investigations into the company including the United Kingdom, Australia, Canada, Nigeria, Kenya, and India”).

17 See, e.g., THE GREAT HACK (Netflix 2019).


Clearly, Cambridge Analytica captured the attention of regulators and the public, particularly regarding the use of data in elections. But as Ira Rubenstein described in *Voter Privacy in the Age of Big Data* in 2014, the application of Big Data, behavioral advertising, and advanced data analytics techniques are not necessarily new in political campaigns. Political parties have been some of the earliest adopters of sophisticated online data analytics dating back to the 2004 Presidential election. Rubenstein observes that major political parties may be the “largest assemblages of personal data in contemporary American life.”

As a result, whether the Voter Privacy Act is signed into law or not, at some point courts will likely consider the constitutionality of regulating the use of personal information in elections. Moreover, given the volume of information now in the public record regarding Cambridge Analytica’s use of computational means to discern, analyze, and manipulate voter’s underlying psychological characteristics, there is an additional question as to whether those techniques could reach a level of scientific efficacy for the court to find them violative of fundamental rights like individual liberty and popular sovereignty.

To that end, the following will review Senator Feinstein’s proposed Voter Privacy Act and potential legal obstacles associated with regulating the use of voters’ personal information in elections, as well as the techniques used by Cambridge Analytica and whether they could change courts’ views of voter privacy in the Digital Age.

II. THE VOTER PRIVACY ACT

The proposed Voter Privacy Act of 2019 is not only the first federal legislation that would regulate the use of voters’ personal information in elections, but its provisions seem intended to shape future litigation regarding voter privacy. Broadly, the bill would amend the Federal Election Campaign Act of 1971 (FECA) to provide voters with legal rights to control the use of their personal information by political entities, which are defined as candidates, political committees, national committees, and political parties under the FECA as well as political organizations under Section 527 of the Internal Revenue Code. The bill also applies to any person using voter

24 Id. at 861.
26 Id.
personal information to carry out regulated campaign activities under the FECA, including public communications, mass mailing, or telephone banking.\textsuperscript{27} Voters’ data ownership rights would be enforced by the Federal Election Commission, including via civil and criminal sanctions.\textsuperscript{28}

A. Findings and Sense of Congress

The bill’s \emph{Congressional Findings} attempt to shape future litigation in two ways: (1) by providing sufficient factual detail to allow a court to find that modern voter targeting techniques are materially different from historic political advertising; and (2) in its role as the representatives of the public, clearly articulating Congress’s assessment of the public’s interest in regulating such activity rather than leaving it to the court.

To support subsequent provisions, the bill provides significant detail and resources on the following factual themes: dramatic growth in Internet usage in the United States; proliferation and depth of data collection by online platforms; advancement of new surveillance techniques such as real-time recording of an individual’s browsing sessions; the ability to infer private information that was never revealed publicly; the incorporation of neurological research into online advertising techniques designed to manipulate users’ “precognitive functions”; the adoption of behavioral advertising and microtargeting techniques by U.S. political entities; and Cambridge Analytica’s use of voter’s personality traits to conduct psychological operations to alter voter behavior.\textsuperscript{29}

The sense of Congress also explicitly states the government’s interest based on the foregoing facts: “It is the sense of Congress that . . . the Federal Government has a compelling interest in protecting voters from surveillance and manipulation;” that the bill is “the most narrowly tailored approach to protecting voters from psychological manipulation;” and that the “Federal Government’s interest would justify additional prohibitions” if necessary.\textsuperscript{30}

Congressional findings are not only relevant for courts in determining legislative judgement,\textsuperscript{31} but also in “advanc[ing] judicial review by identifying the factual authority on which Congress relied.”\textsuperscript{32} In addition, there is at least some indication that the amount of data provided by Congress could be dispositive in a difficult balancing of competing interests.\textsuperscript{33}

\textsuperscript{27} See 52 U.S.C. § 30101(21).
\textsuperscript{28} Voter Privacy Act, \textit{supra} note 21, at § 358; \textit{see also} id. at 354(e)(2).
\textsuperscript{29} \textit{Id.} at § 2.
\textsuperscript{30} \textit{Id.} at § 3.
\textsuperscript{31} United States v. Morrison, 529 U.S. 598 (2000).
\textsuperscript{32} \textit{Id.} at 628 (Souter, J., dissenting).
\textsuperscript{33} \textit{Id.}
B. Rights Provided Under the Act

Similar to the European Union’s General Data Protection Regulation (GDPR)\textsuperscript{34} and the California Consumer Privacy Act (CCPA),\textsuperscript{35} the Voter Privacy Act would allow voters to control the use of their personal information via certain codified rights: (1) right of access, (2) right of erasure, (3) right to prohibit transfer, (4) right to notice, and (5) right to prohibit targeting.\textsuperscript{36} Collectively, these rights allow voters to dictate how their information is used by covered entities, though voters are required to act affirmatively. The only right in the bill that is self-executing is the right of notice, which requires covered entities to notify voters as soon as they take possession of voters’ personal information.

One of the Voter Privacy Act’s novel components is its prohibition on targeting by third-party websites on behalf of covered entities. The provision is similar to Twitter’s recent announcement that it no longer accepts political advertisements on its platform.\textsuperscript{37} As has been the source of some confusion, Twitter’s ban on political advertising does not preclude politicians from using the platform, including for political messaging.\textsuperscript{38} Rather, Twitter will no longer allow political entities to purchase its ad-targeting services that use sophisticated user data profiles to force ads into individuals’ Twitter feeds.\textsuperscript{39}

The Voter Privacy Act’s prohibition on targeting seems to work similarly to Twitter’s policy. It would allow individuals to opt out of permitting third-party websites to use their personal information for ads targeted on behalf of a covered entity. For example, an individual could log onto Facebook, go to their privacy settings, and opt out of receiving targeted ads from covered entities. The provision would not limit covered entities’ ability to send out ads on third-party websites, nor would it limit the total number of online ads a covered entity could send. It would only limit covered entities’ ability to use personal information to target its online ads. Notably, the bill does not include information derived from state voter registration databases in its definition of personal information. As a result, covered entities


\textsuperscript{35} CAL. CIV. CODE § 1798 et seq. (2018).

\textsuperscript{36} Voter Privacy Act, supra note 21, at § 4.


\textsuperscript{38} Id.

\textsuperscript{39} Id.
could still target online ads, even if a voter opts out, based on data like party affiliation, address, and zip code.40

One clear difference between the Voter Privacy Act and other data protection statutes like the GDPR is that the bill does not generally include front-end “use restrictions.”41 The GDPR limits covered entities to six exclusive, authorized uses for individual personal information. The authorized uses describe general categories of activity like consent that is necessary for the performance of a contract or necessary for performance of a legal duty.42 Any collection, use, or processing of personal information inconsistent with an approved use is prohibited. The Voter Privacy Act has no such restriction. Political entities are not subject to any blanket prohibition on collection, use, or processing of any voter’s data. The only limitation a covered entity would encounter would be from the voter themselves, and only with respect to that voter’s own personal data.

The lone use restriction on data transfer in the legislation is a prohibition on transferring voters’ personal information outside of the United States.43 There is no specific explanation for the provision. However, given that Senator Feinstein indicated on numerous occasions that the bill was intended to respond to Cambridge Analytica, one possible reason for the provision could be a response to reports that Cambridge Analytica passed U.S. voter information to the state-owned Russian oil firm Lukoil.44 At the Senate Judiciary Committee’s hearing on Cambridge Analytica, Senator Feinstein referred to Lukoil as having a “formal information-sharing partnership with the Russian Federal Secret Service, the FSB, the successor to the KGB.”45

C. Definitions Included in the Act

The Voter Privacy Act’s definitions provide notable limitations on the overall scope of the legislation. For example, the bill defines “personal information”46 to include ten categories of data like personal identifiers, consumer history, Internet browsing history, geolocation, personality traits, and other psychographic modeling.47 However, it excludes publicly available

---

40 Voter Privacy Act, supra note 21, at § 351(4)(B).
41 General Data Protection Regulation, supra note 34, at Art. 6.
42 Specifically, these authorized uses include: (1) use with consent, (2) use in performance of a contract, (3) use in compliance with another legal obligation, (4) use to protect the vital interests of the data subject, (5) use for the public interests, and (6) use for the legitimate interests of the controller. Id.
43 Voter Privacy Act, supra note 21, at § 4.
44 Cadwalladr & Graham-Harrison, supra note 1.
45 Feinstein, supra note 16.
46 Voter Privacy Act, supra note 21, at § 4(a).
47 Id.
information, deidentified information, and aggregate polling information.\textsuperscript{48} Publicly available information is defined as “information obtained from a Federal, state, or local voter registration database.”\textsuperscript{49}

The exclusion of information obtained from state voter registration databases is significant for two reasons. One, according to the National Conference of State Legislatures, state voter registration databases can include significant information including party registration, name, address, date of birth, occupation, and voting history.\textsuperscript{50} Two, it ensures that candidates have access to some minimum voter data to protect their ability to effectively communicate with voters. As Rubenstein observes, some political professionals consider state voter registration data the most valuable data available in operating campaigns.\textsuperscript{51}

III. LEGAL CHALLENGES

Endeavoring to regulate the use of personal data in elections, the Voter Privacy Act could face a number of constitutional challenges. The most likely are: (A) whether the law impermissibly limits the quantity of protected speech;\textsuperscript{52} and (B) whether the law impermissibly restrains the individual right of candidates to access a voter’s personal information.\textsuperscript{53}

A. Limits on Quantity of Speech

The first test for any legislation that regulates political campaign activity, particularly under the Federal Election Campaign Act (FECA), is under \textit{Buckley v. Valeo}. Following the Watergate scandal and subsequent passage of the Federal Election Campaign Act of 1974 (FECA),\textsuperscript{54} plaintiffs in \textit{Buckley} challenged a variety of the new campaign finance laws, including expenditure limits, contribution limits, public disclosure of contributions, and public financing of elections.\textsuperscript{55} Expenditure limits in the Act capped the overall amount a candidate could spend during a campaign and how much an

\textsuperscript{48} Id.
\textsuperscript{49} Id.
\textsuperscript{50} Rubenstein, \textit{supra} note 23, at 868.
\textsuperscript{51} Id.
\textsuperscript{52} See \textit{Buckley v. Valeo}, 424 U.S. 1 (1976).
\textsuperscript{54} Lee E. Goodman, \textit{The First Amendment Right to Political Privacy, Chapter 6, Campaign Finance and Other Very Public Exceptions to Privacy}, WILEY REIN LLP (Sept. 2019), https://www.wiley.law/newsletter-The-First-Amendment-Right-to-Political-Privacy-Chapter-6-Campaign-Finance-and-Other-Very-Public-Exceptions-to-Privacy [https://perma.cc/5STJ-QFSD].
\textsuperscript{55} \textit{Buckley}, 424 U.S. at 19–20.
individual could spend “relative to a clearly identified candidate.”\textsuperscript{56} The Court invalidated limits on expenditures finding that “a restriction on the amount of money a person or group can spend on political communication during a campaign necessarily reduces the quantity of expression by restricting the number of issues discussed, the depth of the exploration, and the size of the audience reached.”\textsuperscript{57}

The Court upheld other provisions of the FECA such as mandating disclosure of political contributions. The FECA required political committees to keep detailed records of all contributions and expenditures, including the name and address of every individual contributor over a certain dollar amount, and to publish them for public inspection. The Court found that while the provision imposed no ceiling on campaign related activities, it compelled disclosure of political activity and infringed on the “privacy of association and belief guaranteed by the First Amendment.”\textsuperscript{58} However, potentially in part because it had already invalidated expenditure limits, the Court conceded that disclosure was the least restrictive means for Congress to address the rampant corruption it had identified in political campaigns and upheld the provision.

Over the next forty years the Court would take an increasingly aggressive posture toward campaign finance regulations, invalidating provisions like limits on independent expenditures by corporations.\textsuperscript{59} Yet, it has remained equally steadfast in upholding public disclosure requirements.\textsuperscript{60} Outside of right-of-association cases, the only instances where the Court has retreated from its preference for disclosure over privacy were in instances where plaintiffs could demonstrate that compelled disclosure had actually resulted in physical harm.\textsuperscript{61}

\textit{Buckley} and its progeny continue to stand for the proposition that any regulation under the FECA may not restrict the \textit{quantity} of speech. On its face the Voter Privacy Act does not appear to run afoul of that requirements. However, the Court’s repeated preference for transparency at the expense of privacy is instructive on the relative weight of two interests in the context of political campaigns.

B. Limits on Candidates’ Access to Data

The Voter Privacy Act could limit a candidate’s access to certain voter data, which the Court has found is protected speech under the First

\textsuperscript{56} Id. at 13.
\textsuperscript{57} Id. at 19.
\textsuperscript{58} Id. at 64.
\textsuperscript{60} See McConnell v. FEC, 540 U.S. 93 (2003).
\textsuperscript{61} Britt v. Superior Court, 20 Cal. 3d 844, 849 (1978).
Amendment. In *Sorrell v. IMS Health Inc.*, the Court invalidated a Vermont statute that prohibited the sale of prescriber-identifiable information for marketing prescription drugs without the prescriber’s consent. The Court found that, by prohibiting access to data for a singular purpose like marketing, the statute imposed content-based, viewpoint-based, and speaker-based restrictions under the First Amendment.

Finding that the regulation imposed significant restrictions on protected speech, the Court considered whether the government had articulated a substantial interest sufficient to sustain the burden. Vermont argued that the statute was necessary to protect medical privacy and to achieve several policy objectives, like improving public health and reducing health care costs. The Court rejected both arguments, finding while it restricted access to marketers, that the Vermont statute permitted access to journalists, researchers, and insurers. The Court also found that while the state’s goals to reduce health care costs “may be proper, [the statute] does not advance them in a permissible way,” namely, by “restraining certain speech by certain speakers.”

*Sorrell* continues to stand for the proposition that access to data, which can be used to improve the persuasiveness of a speaker’s message, is protected under the First Amendment. Notably, Justice Kennedy did not foreclose the possibility of any regulation of personal data under the First Amendment. In his opinion, he cites data protection statutes like Health Insurance Portability and Accountability Act (HIPAA) as positive examples that did not raise the same concerns as the statute at issue. However, *Sorrell* does continue to stand for the proposition that if a statute endeavors to regulate the use of data, it must serve a sufficiently compelling interest.

IV. IS VOTER PRIVACY A COMPPELLING INTEREST IN THE DIGITAL AGE?

Whether voter privacy could be considered a sufficiently compelling interest to sustain legislation like the Voter Privacy Act is a two-part question. The first part is a factual question about whether there is sufficient data or other evidence to find a concept like voter privacy compelling on its face. The

---

63 *Id.* at 557.
64 *Id.* at 564–65.
65 *Id.* at 572 (“First, the State contends that its law is necessary to protect medical privacy, including physician confidentiality, avoidance of harassment, and the integrity of the doctor-patient relationship. Second the State argues that §4631(d) is integral to the achievement of policy objectives—namely, improved public health and reduced health care costs”).
66 *Id.* at 572–73.
67 *Id.* at 577.
68 *Id.* at 573.
second could be described as more of a question of law, but essentially it is how a court would choose to balance voter privacy against other competing rights and interests.

A. Factual Basis for Voter Privacy

Fortunately, investigations of Cambridge Analytica have provided a wealth of information regarding modern data practices by political campaigns, advances in psychographics, and the national security implications of foreign nations adopting these techniques to engineer outcomes in American elections.

1. Modern Data Collection and Campaigns

The Internet is an indispensable part of modern economic and social life in the United States. According to the Pew Research Center, ninety percent of Americans use the Internet. From 2005 to 2019, the number of Americans using social media increased from five percent to seventy-two percent. For those under the age of thirty, that figure was ninety percent. Facebook has 2.4 billion daily active users globally. About seventy percent of American adults have a Facebook account, and forty-three percent get their news from Facebook.

As a result, the repository of personal information that is collected and made widely available by companies like Google and Facebook about each individual American citizen is materially different than any period in American history. As noted in the Voter Privacy Act:

One U.S. based search engine advertises its ability to track hundreds of categories of data about specific individuals including age, gender, occupation, income level, sexual orientation, national origin, religion, medical conditions such as AIDS, erectile dysfunction, bipolar disorder, eating disorders, and sexually transmitted diseases, family

---

70 Id.
71 Id.
information such as number of children, children with special needs, infertility, and substance misuse, and support for social issues such as reproductive rights, unions and labor issues, and support for gun rights.\textsuperscript{74}

Every time an individual visits a website, information like the kind listed above is broadcast to tens or hundreds of companies for the opportunity to show that individual advertisements.\textsuperscript{75} Moreover, these advertising protocols include unique ID codes “so that all of this information can be tied to you over time. This allows companies you have never heard of to maintain intimate profiles on you, and on everyone you have ever known.”\textsuperscript{76} One reporter determined that Google maintained 5.5 gigabytes of her personal data, which is roughly equivalent to 3 million word documents worth of text.\textsuperscript{77}

Campaigns have been some of the earliest adopters of online data collection, analytics, and microtargeting. As early as 2004, major party databases included data on “every one of the 168 million or so registered voters in the country, cross-indexed with phone numbers, addresses, voting history, income range and so-on up to several hundred points of data on each voter.”\textsuperscript{78} In 2014, Rubenstein observed:

Political databases hold records on almost 200 million eligible American voters. Each record contains hundreds if not thousands of fields derived from voter rolls, donor and response data, campaign web data, and consumer and other data obtained from data brokers, all of which is combined into giant assemblages made possible by fast computers, speedy network connections, cheap data storage, and ample financial and technical resources.\textsuperscript{79}

Unique personalized identifiers like IP addresses, cookies, and mobile device IDs allow campaigns to integrate diverse datasets “while data mining and sophisticated statistical techniques allow them to engage in highly strategic and cost-effective analysis and targeting.”\textsuperscript{80} Because political professionals consider data dispositive in determining the outcome of

\textsuperscript{74} Voter Privacy Act, supra note 21, at § 2(3).
\textsuperscript{75} Id., at § 5.
\textsuperscript{76} Id.
\textsuperscript{77} Dylan Curran, Are You Ready? Here is All the Data Facebook and Google Have on You, GUARDIAN (Mar. 30, 2018), https://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy [https://perma.cc/PT5X-N58D].
\textsuperscript{78} Rubenstein, supra note 23, at 876.
\textsuperscript{79} Id. at 879.
\textsuperscript{80} Id. at 884.
elections, campaigns will continue to expand the scope of their collection, analysis, and targeting.

2. New Developments in Psychographics

Cambridge Analytica used large datasets and computational means to target and influence voters’ underlying psychological traits to alter their behavior. The firm reportedly based its work on a series of studies published by the National Academy of Sciences that describe material advances in the field of psychographics made possible by U.S. social media platforms.81

Psychographics is the study and classification of people according to their attitudes, aspirations, and other psychological criteria. Whereas *demographics* measure age, education, income, gender, and race, *psychographics* measure psychological traits like personality, interests, attitudes, and beliefs. Rather than describing who a person is, psychographics attempt to illustrate why a person behaves a certain way. Psychographics were developed after World War II as researchers sought to apply concepts of clinical psychology in order to better understand consumer behavior.82 As one scholar described, psychographic information puts flesh on the demographic bone.83

One of the most common psychographic measures is personality. Like most psychographics, it is a latent trait—meaning it cannot be observed directly. Instead, individual characteristics are measured through surveys and questionnaires. One of the most common methods for testing personality is the Five-Factor Model, also called the “OCEAN test.” The OCEAN test categorizes an individual’s personality according to the following traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism.

It is well-established that persuasive appeals are more effective in influencing human behavior when they are tailored to an individual’s unique personality traits.84 However, large-scale personality-based persuasion campaigns face two limitations: cost and reliability. It would not be practical or cost effective to attempt to convince five to ten million people to sit for a one-hundred question personality test. Alternatively, focus groups can yield useful insight, but the reliability of the results can degrade with extrapolation to larger groups of diverse individuals.85

83 *Id.* at 198.
85 See Wells, *supra* note 82.
Cambridge Analytica reportedly based its methodology on two recent studies published in *Proceedings of the National Academy of Sciences*, the official journal of the National Academy of Sciences. The studies suggest that personality traits can be cheaply and accurately inferred from an individual’s public social media data, which can be used to conduct large-scale psychologically persuasive campaigns based on actual knowledge of each target’s personality rather than estimations.

The first study found that individual’s personality traits can be measured through social media data, instead of requiring a personality questionnaire. Entitled *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, the study examined 58,000 volunteers, each of whom provided Facebook Likes, demographic profiles, and the results of several personality questionnaires. The resulting data was analyzed to connect Facebook Likes for topics such as “the Colbert Report,” “Harley Davidson,” or “Wu Tang Clan” with personality traits like “openness,” as well as other attributes like sexual orientation, political views, or substance abuse. Researchers concluded “basic digital records of human behavior can be used to automatically and accurately estimate a wide-range of personal attributes that people would typically assume to be private... [P]redictability of individual attributes from digital records of behavior may have considerable negative implications, because it can easily be applied to large numbers of people without obtaining their individual consent and without them noticing.” Subsequent studies have found success at inferring sensitive personal information from personal websites, blogs, Twitter messages, Facebook profiles, and Instagram pictures. Facebook itself reportedly

---


87 Id.

88 Id. at 5894.

89 Id. at 5805.


91 Tal Yarkoni, *Personality in 100,000 Words: A Large-Scale Analysis of Personality and Word Use Among Bloggers*, 44 J. RES. PERSONALITY 363, 363–373 (2010).


obtained a patent that described how personality characteristics like emotional stability could be determined from individuals’ messages and status updates.95

The second study found that computers could infer an individual’s personality traits more accurately than any human. Entitled Computer-Based Personality Judgements Are More Accurate Than Those Made by Humans, the study compared computer projections of an individual’s personality with those made by the same individual’s coworkers, friends, family members and spouse.96 The study concluded that computers were “significantly more accurate than humans” at predicting an individual’s personality.97 Moreover, with only ten, seventy, one hundred and fifty, and three hundred Facebook Likes, computers could outperform an average coworker, friend, family member, and spouse respectively.98 Researchers noted that that “growth in both the sophistication of the computer models and the amount of digital footprint might lead to computer models outperforming humans even more decisively.”99 Again, researchers warned that:

[A]utomated, accurate, and cheap personality assessment tools could affect society in many ways ... knowledge of people’s personalities can also be used to manipulate and influence them. Understandably, people might distrust or reject digital technologies after realizing that their government, internet provider, web browser, online social network, or search engine can infer their personal characteristics more accurately than their closest family members.100

Together, the studies addressed two of the fundamental challenges of applying large-scale psychographic persuasion campaigns. First, researchers can infer the personality traits of large populations of individuals for minimal cost and without those individuals’ knowledge. Second, because computer algorithms can determine the personality for each individual in a large population—rather than by generalizing from focus groups and modeling—the results can be highly accurate.

In 2017, a third study found that using inferred personality data from social media allowed researchers to alter the behavior of large groups of

---

96 Youyou et al., supra note 11, at 1037.
97 Id.
99 Youyou et al., supra note 11, at 1039.
100 Id.
people in a real-world environment. Entitled *Psychological Targeting as an Effective Approach to Digital Mass Persuasion*, the study conducted three real-world experiments including 3.7 million people, targeting each individual based on their unique openness and extraversion qualities as derived from Facebook Likes. Even with a small digital footprint for each individual, matching psychologically persuasive ads to each individual’s personality traits resulted in 40% more clicks and 50% more purchases. Researchers concluded:

The results of these three studies provide converging evidence for the effectiveness of psychological targeting in the context of real-life digital mass persuasion; tailoring persuasive appeals to the psychological profiles of large groups of people allowed us to influence their actual behaviors and choices. Given that we approximated people’s psychological profiles using a single [Facebook] Like per person—instead of predicting individuals profiles using people’s full history of digital footprints—our findings represent a conservative estimate of the potential effectiveness of psychological mass persuasion.

Further, researchers commented that these types of psychologically persuasive advertisements could be applied by governments, companies, or political parties “to covertly exploit weaknesses in [peoples’] character and persuade them to take action against their own best interest, [highlighting] the potential need for policy interventions.”

There is sufficient evidence to presume that techniques like digital mass persuasion can be effective in highly networked societies with large repositories of personal data. The key to both the accuracy and effect of these methodologies is the amount of data available on each subject. It is well-established that personal data is a strategic economic asset. Given that personal data could also be used to engineer the outcome of an election, democracies might begin to consider it a national security asset as well.

---

101 Matz et al., *supra* note 84.
102 *Id.* at 12717.
103 *Id.* at 12714.
3. National Security

Given the Russian government’s recent use of information operations to undermine American elections and their effort to obtain U.S. voter data from Cambridge Analytica directly, the U.S. government has a clear national security interest in limiting foreign access to U.S. voters’ personal information.

It is well established that the Russian government conducted large-scale information operations targeting U.S. voters to help elect Donald Trump as U.S. President. On January 6, 2017, the Director of National Intelligence (DNI) released a consensus Intelligence Community assessment regarding Russian interference in the 2016 U.S. Presidential election. The report concluded that Russia used a combination of government agencies, state-funded media, third-party intermediaries, and paid social media users or trolls to undermine Americans’ confidence in democracy and help elect Donald Trump. The DNI concluded that the effort was a part of “Moscow’s longstanding desire to undermine the U.S.-led liberal democratic order.”

In March 2019, Special Counsel Robert S. Mueller’s Report on the Investigation into Russian Interference in the 2016 Presidential Election found that the Russian government used U.S. social media platforms extensively to influence the 2016 election in favor of Donald Trump. Facebook testified that 470 Facebook accounts controlled by the Russian Internet Research Agency (IRA) made 80,000 posts between 2015 and 2017 reaching potentially 126 million persons. In addition, “to reach larger audiences, the IRA purchased advertisements from Facebook that promoted IRA groups on the newsfeeds of U.S. audience members.” Using paid advertising tools on social media allowed Russian operatives to access Facebook’s sophisticated ad-targeting services built on large repositories of American’s sensitive personal information.

106 Id. at ii.
108 The Internet Research Agency is an organization of professional Internet trolls whose activities are coordinated by the Russian government. See id. at 4.
109 Id. at 15 (citing Social Media Influence in the 2016 U.S. Election: Hearing Before the Senate Select Committee on Intelligence, 115th Cong. 13 (Nov. 1, 2017) (testimony of Colin Stretch, General Counsel, Facebook)).
110 Id. at 25.
At the same time that Russian government was targeting American voters through U.S. social media platforms, Cambridge Analytica reportedly passed information to Russian intelligence services on its U.S. voter targeting activities via the state-owned Russian oil firm Lukoil. According to Senator Feinstein, Lukoil has a formal information-sharing agreement with the Russian Federal Security Service (FSB), which is the successor to the KGB. In addition, Cambridge Analytica used a Moldovan-born professor to extract 87 million Americans’ Facebook data, primarily the results of an online personality questionnaire, that would form the basis of the firms psychological voter-targeting program. Although the professor was associated with Cambridge University in London, he reportedly concealed the fact that he was also receiving compensation from St. Petersburg University in Russia. U.K. Parliament Member Damien Collins, who led the Parliament’s investigation into Cambridge Analytica, also confirmed that the same professor’s data on U.S. voters was accessed from Russia.

Authoritarian governments use U.S.-based social media platforms to conduct information operations not only in the United States, but also all over the world. Reporting suggests that Russia is using the same targeted misinformation on Facebook and Twitter in Africa, the Middle East,
Latin America,\textsuperscript{118} and the Baltic States.\textsuperscript{119} The Senate Foreign Relations Committee found that Russian “disinformation efforts can now take advantage of increasingly powerful analytics that identify ‘customer sentiment,’ allowing them to target the most susceptible and vulnerable audiences.”\textsuperscript{120} The Carnegie Endowment for Peace found that prior to disbanding, Cambridge Analytica “ha[d] been active in Mexico, Brazil, and possibly Columbia, raising the specter of Russian or other external attempts to manipulate public opinion during an election year. Tipping just one or two countries toward an anti-U.S. stance—especially long-time U.S. partners—could complicate U.S. policy and distract Washington from its global priorities.”\textsuperscript{121} Finally, the Computational Research Project at Oxford University found evidence of social media manipulation campaigns in 70 countries in 2019—up from 23 in 2017; these campaigns were conducted via computational propaganda intended to shape domestic public opinion by governments or political parties.\textsuperscript{122}

Cambridge Analytica illustrated the Russian government’s direct interest in U.S. voter data. Because the Russian government sought such data—while conducting information operations targeting U.S. voters that closely resembled Cambridge Analytica’s services to a U.S. Presidential candidate—there is additional evidence that Cambridge’s data and or methodology achieved some level of effectiveness. That the Russian government has continued to expand the same type of operations around the world further speaks to the efficacy of these tactics.

\textsuperscript{118} Julia Gurganus, \textit{Russia: Playing a Geopolitical Game in Latin America}, CARNEGIE ENDOWMENT FOR INT’L PEACE (May 3, 2018), https://carnegieendowment.org/2018/05/03/russia-playing-geopolitical-game-in-latin-america-pub-76228 [https://perma.cc/W2FG-KVHD] (“Populations in Latin America are particularly avid social media users compared to other regions in the world, according to data collected in 2015, making them susceptible to potential Russian efforts to promote divisive or anti-U.S. narratives via online platforms.”).


\textsuperscript{120} STAFF OF THE S. COMM. ON FOREIGN RELATIONS, 115TH CONG., PUTIN’S ASYMMETRIC ASSAULT ON DEMOCRACY IN RUSSIA AND EUROPE: IMPLICATIONS FOR U.S. NATIONAL SECURITY 2 (Comm. Print 2018).

\textsuperscript{121} Gurganus, supra note 118..

Given the advances in the efficacy of mass persuasion techniques, powered by U.S. social media platforms and their large repositories of personal data, information operations ought to be considered a strategic level concern. They have the potential to undermine elections in the United States and in partner democracies around the world, particularly those with less developed civil societies. The Cambridge Analytica affair demonstrated that hostile nations are actively seeking sensitive U.S. voter information. As a result, the government has a clear interest in some regulation of the collection, use, and transfer of voters’ personal information.

B. Rights Balancing and the Legal Argument

Even if a court were to find the factual basis supporting voter privacy compelling, that court would still be required to balance those interests against other competing rights and whether any restrictions were properly tailored to meet those interests.123 To that end, the Voter Privacy Act provides a useful test case to consider the potential competing interests associated with regulating the use of voters’ data as described in Section A, namely voters’ privacy versus transparency and speech.

1. The Voter, the Candidate, and the People

Although courts have generally favored transparency over privacy with respect to regulating political campaigns, the Voter Privacy Act could reorient the equities underlying those decisions. In the context of campaign finance, courts have generally understood privacy as an individual right, and transparency or anticorruption as a right of the public or “the People.” In Buckley, when describing privacy the Court uses terminology associated with the individual: “[disclosures] will deter some individuals who might otherwise contribute,”124 “disclosures may even expose contributors to harassment or retaliation,”125 and “not insignificant burdens on individual rights.”126 In contrast, when describing transparency interests, the Court uses terminology emphasizing the collective: “Disclosure provides the electorate with

123 See Sorrell v. IMS Health, Inc., 564 U.S. 552, 577 (2011) (explaining that Vermont’s chosen manner of restricting speech is too burdensome); see also United States v. Stevens, 559 U.S. 460, 482 (2010) (holding that a content-based speech restriction was overbroad and so not properly tailored).
124 See Buckley v. Valeo, 424 U.S. 1, 96 (1976) (Burger, J., concurring in part) (emphasis added) (citation omitted).
125 Id. at 69 (emphasis added).
126 Id. (emphasis added).
information,“127 exposes [corruption] to the light of publicity,”128 and “a public armed with information about a candidate’s most generous supporters.”129

Implicit in the Court’s reasoning in *Buckley* is a twofold analysis. The first is a normative judgement about transparency and privacy in elections. The second is a utilitarian balancing of the interests of the many in transparency versus the interests of the few in privacy. The Court’s reasoning here appears sound; elections are about effectuating the will of the electorate.

However, the same logic yields a different result as it relates to the Voter Privacy Act. First, whereas in *Buckley* the Court considered two constituents—public and individual—the Voter Privacy Act appears to have three—voter, candidate, and the public. Classifying the interests associated with the first two is relatively straightforward. The voter has a privacy interest in being able to control the use of their personal information in an election, including to mitigate the increasing potential for psychological manipulation by organizations like Cambridge Analytica. The candidate has a speech interest in access to voters’ personal information in order to communicate more persuasively with voters.

Because the Voter Privacy Act allows voters to dictate the use of their own personal information once in possession of a candidate, the inevitable conflict and the Court’s analysis would likely begin with those two competing claims—voter and candidate. Unfortunately, there is no definitive instruction in *Buckley* about the relative weight of an individual voter’s privacy interest versus a candidate’s speech interest.

It is clear from *Sorrell* that limiting access to data is a burden on protected speech. Therefore, ordinarily whether the Court would sustain such a regulation would turn on the government’s countervailing interest. However, in the case of the Voter Privacy Act, the competing interest is not the government’s—it’s the voter’s. In the context of an election, which is a direct exercise of sovereign authority in democracy, it seems likely the Court would find the voter’s interest more compelling. As described in Section A, the voter’s interest is even more compelling in light of the information popularized by Cambridge Analytica: the more data that is known about a voter, the greater the likelihood of a third-party altering that voter’s behavior through psychological manipulation. However, the simplest way to think about it might be: it would be an odd holding for the Court to find that a candidate seeking representative office has more of a right to voters’ personal information than the voters themselves.

---

127 Id. at 66 (emphasis added).
128 Id. at 67 (emphasis added).
129 Id. (emphasis added).
The third interest the Court would likely consider is the public or the People. In *Buckley*, the public interest was rightly aligned with transparency. The Court reasoned that though some wealthy campaign contributors would have to disclose their political contributions, the vast majority of the electorate would benefit from elections less likely to include corruption. Again, the Voter Privacy Act seems to alter that assessment.

As Dr. Priscilla Regan rightly points out, political privacy is not only an individual right, but a “public value that supports democratic political systems.” Therefore, the privacy necessary for an individual voter to exercise independent judgement in an election is not only an interest of a voter, but of all voters—and even democracy itself.

In addition, Cambridge Analytica’s connection between the amount of data available about a voter and potential for psychological coercion further supports the public’s interest in voter privacy. First, the public has an interest in being governed by a government that is a legitimate reflection of the People’s will. Second, the public has a national security interest in mitigating the ability of foreign nations to manipulate the electorate through mass digital persuasion like the kind implemented by the Russian government, and potentially aided by Cambridge Analytica. On the other hand, there does not appear to be the same persuasive case that the public has an interest in a candidate’s having unregulated access to voters’ personal information, particularly over the objection of the voters themselves.

Therefore, even if the Court did not find that an individual voter’s right to privacy outweighed a candidate’s right to access that voter’s personal information, the combination of the voter and the public’s interest would be dispositive.

2. *Preserving a Candidate’s Access to Data*

The second consideration is whether the Voter Privacy Act’s potential limitation on access to data would undermine a candidate’s ability to communicate persuasively with the electorate. In *Sorrell*, the court confirmed that access to data is protected speech and that salespersons are more effective when they know the background and purchasing preferences of their clientele. It is also well-established that the Court’s concerns would be heightened in the context of political speech.

---


131 Sorrell v. IMS Health, Inc., 564 U.S. 552, 571 (2011) (“Facts, after all, are the beginning point for much of the speech that is most essential to … conduct human affairs”).

132 Id.
The Voter Privacy Act includes a number of important limitations that could mitigate those concerns. Unlike the statute at issue in *Sorrell*, the Voter Privacy Act does not include an initial restriction on any voter’s data. Candidates could continue to collect any type of voter data and would continue to have access to that data, absent a voter’s intervention.

The Voter Privacy Act also excludes three categories of data from voter control: publicly available information, deidentified information, and aggregate polling information. Meaning, if a candidate only obtained information from state voter registration databases and anonymized polling data, she would not be subject to any voter instructions or limitations on the use of that data because it is outside the scope of the bill.

Moreover, the data contained in voter registration databases appears to be meaningful. They include significant information like name, address, signature, date of birth, phone number, gender, party affiliation, and voting history.\(^\text{133}\) And, there is evidence that it is some of the most useful information available. According to Rubenstein, state voter registration databases “play[] a critical role in the U.S. political system by enabling candidates and others to communicate with voters for political purposes by mail, phone, email, and door-to-door canvassing.”\(^\text{134}\) Moreover, at least some academics and political professionals argue that, as of 2014, publicly-available voter registration data was the most important source of information for campaigns.\(^\text{135}\)

The absence of blanket restrictions on voters’ data coupled with the exclusion of state voter registration databases and anonymized polling data from the bill ensures candidates would have access to at least some meaningful voter data. Defenders of the legislation would likely point to these provisions as evidence that the bill is narrowly tailored; however, it is likely to also factor in the court’s balancing of rights as well.

V. CONCLUSION

There may actually be more than one silver lining associated with Cambridge Analytica. The first is that this affair firmly established the challenges associated with the application of Big Data to democratic elections in the public conscience. The firm provided a road map for citizens and policy makers to consider the emerging science around Big Data and the national security risks associated with unregulated access to voters’ personal information.

And, with credit to Senator Feinstein, the affair also resulted in the introduction of the first federal legislation in our history that would regulate

---

134 *Id.*
135 *Id.* at 886.
the use of voters’ personal data in elections. Hopefully, it will also result in courts reconsidering concepts like voter and political privacy and whether they merit increased attention in the Digital Age.
PUBLIC SAFETY AND DISINFORMATION

Karen Kornbluh* & Eli Weiner**

CITE AS: 4 GEO. L. TECH. REV. 609 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 610
II. THE PROBLEM ........................................................................ 611
   A. Misleading Outlets, Hyper-Partisan Clickbait, and Arbitraging
      The Trust Built Up by Independent Journalism ....................... 611
   B. Personalized Political Propaganda Obscures The True Sponsors
      of Online Ads .................................................................. 612
   C. Paid Influencers and Networks of Amplifiers Have Become
      Critical in 2020 ................................................................. 614
   D. Secret Groups, Encrypted Messaging, and Fringe Sites Linking to
      the Main Platforms .......................................................... 616
   E. Loopholes Created by Different Platform Rules Are Applied
      Inconsistently and Without Transparency, Frustrating Accountability. 617
III. NEW DIGITAL DEMOCRACY ARCHITECTURE ..................... 620
   A. Update Offline Protections for an Online World ...................... 621
   B. Provide Choice by Funding a PBS of the Internet Through a
      “Superfund” Type Fee on Digital Advertising Revenue ............. 622
   C. Create Accountability With a Code of Conduct Enforceable
      Through a Combination of the Following: ............................ 622
IV. CONCLUSION ......................................................................... 623

* Senior Fellow and Director, Digital Innovation and Democracy Initiative, German Marshall Fund of the United States. This article is a precis of Karen Kornbluh and Ellen P. Goodman, Safeguarding Digital Democracy Digital Innovation and Democracy Roadmap, GERMAN MARSHALL FUND OF THE UNITED STATES (Mar. 24 2020), http://www.gmfus.org/publications/safeguarding-democracy-against-disinformation [https://perma.cc/B4DN]. I am deeply thankful to the Democracy Fund, the Hewlett Foundation, and the Knight Foundation for supporting the Digital Innovation and Democracy Initiative at the German Marshall Fund of the U.S.; Rachel Tausenfreund who edited the original report; research assistant Ryan Whittington; and the Georgetown Law Technology Review for their helpful edits. Any mistakes herein are my own.

** Eli Weiner is a research assistant with the Digital Innovation and Democracy Initiative at the German Marshall Fund of the United States.
I. INTRODUCTION

Too often, combatting public safety and disinformation on the Internet is presented as a false choice between continuing to allow platforms free reign to set rules of the road for our digital media ecosystem and giving the government more control over the content flowing across the networks. Currently, the government in India is moving to clamp down on Internet companies with sweeping new regulations that would force them to take down any content deemed “unlawful in any manner whatever.” Furthermore, in the United States Senate, legislative proposals would have government agencies condition immunity on whether Internet companies exhibit political bias or approve platform best practices.

Other options would empower users instead of allowing either platforms or government to act as a censor. With an understanding of the digital information platforms themselves, new media gatekeepers could suggest options that would update the obligations of our old gatekeepers—to minimize the opportunity for user manipulation, boost public interest journalism, and promote democratic debate. A new media architecture would steer clear of vague rules and instead focus on updating offline protections, fostering choice, and public accountability. It would be technology-neutral and tailored with input from stakeholders. Moreover, in so doing, it would close the loopholes that allow bad actors to engage in online information warfare on the largest platforms without restricting free expression or stymieing welcome innovation.

At the German Marshall Fund of the United States, we have tackled these issues in a more detailed report. The new architecture would have three major elements: (1) increasing offline rights and protections for consumers, elections, civil rights, and privacy; (2) promoting and sustaining local public news and media literacy; and (3) creating an accountability structure for these elements and content moderation.

II. THE PROBLEM

“On the Internet, no one knows you’re a dog,” reads the caption from the famous New Yorker cartoon of a dog at a computer. While personal anonymity has always been a feature of the Internet—until Facebook’s policy of verifying accounts and Twitter’s blue checks—a host of newer design features have developed which allow parasitic campaigns to manipulate users on the platforms, weaponizing tribal fears and corrupting the information ecosystem with disinformation. Partisan and clickbait outlets have the same design features as independent journalistic outlets, though they adhere to none of the standards of journalism (e.g., the masthead, corrections, separation of news from opinion). Online ads leverage users’ data usually without their understanding to test and target content to the most susceptible; even so-called “organic content” (activity by other users) can be the hidden result of bots or coordinated activity. There are five major categories of online activity through which disinformation campaigns can deceive users: (A) misleading outlets; (B) personalized political propaganda; (C) paid influencers and networks of amplifiers; (D) secret groups, encrypted messaging, and fringe sites linking to main platforms; and (E) inconsistently-applied loopholes created by different platform rules.

A. Misleading Outlets, Hyper-Partisan Clickbait, and Arbitraging The Trust Built Up by Independent Journalism

The user interfaces used by platforms ensure the appearance of their stories is the same as those from traditional journalistic organizations, while separating the stories from the outlet that produces it. This interface denies the reader access to information developed by independent journalism to offer readers transparency about the news sources (including bylines, mastheads, separated news from opinion, corrections, and codes and standards)—or this interface can even obscure whether the news outlet provides this information at all. Readers who see some stories with fact-checks may assume that stories without those have been checked and deemed factual, when in fact, only a portion are fact-checked, and stories bear no indication if they are satire or opinion. Not only do these practices boost trust in disinformation, they also, over time, undermine trust in all news.

Meanwhile, legacy newspapers struggle as platforms have continued to cannibalize the revenue of local independent journalism. Google and Facebook now capture 58 percent of the advertising market.\(^5\) One of the nation’s leading newspaper companies with 30 newspapers around the country, McClatchy, recently filed for bankruptcy.\(^6\) Since 2004, more than 1,800 local print outlets have closed, and at least 200 counties have no newspaper at all.\(^7\)

These statistics imply a civic emergency. Areas with limited local news have less politically aware populations.\(^8\) One study found that the city of Denver experienced a decrease in civic engagement after the closure of *The Rocky Mountain News* and the shrinking of the *Denver Post*.\(^9\) Layoffs and closings have hamstrung the ability to hold public and corporate officials accountable.\(^10\)

**B. Personalized Political Propaganda Obscures The True Sponsors of Online Ads**

Disinformation campaigns advertise to small audiences of users enticing them to share memes, take quizzes, donate, follow “news” sites and fictitious accounts, and to join groups. The ads are targeted to audiences based on data gathered about them and people like them. For example, in one effort, women over the age of twenty-five who had expressed interest in pregnancy were served a targeted ad featuring anti-vaccination conspiracies.\(^11\)

---


\(^8\) *Id.* at 14.

\(^9\) *Id.* at 14.


Although Facebook now “prohibits ads that include claims debunked by third-party fact checkers or, in certain circumstances, claims debunked by organizations with particular expertise,” it decided to exempt ads from political candidates from fact-checking requirements on the grounds that it was important to allow the ads to be subject to public scrutiny.\(^\text{12}\) However, as hundreds of Facebook employees warned in an open letter voicing their objection to the policy, “it’s hard for people in the electorate to participate in the ‘public scrutiny’ that we’re saying comes along with political speech,” because these ads are only shown to small groups.\(^\text{13}\) Google’s ad policies prohibit misleading content, and the company announced that it will only restrict misinformation in political ads that “could significantly undermine participation or trust in an electoral or democratic process,” suggesting that “misleading content” will be defined narrowly to exclude misinformation about specific candidates or policies.\(^\text{14}\) The discrepancy among platform rules creates loopholes that cross-platform disinformation campaigns exploit.\(^\text{15}\)

Even if ads do not contain falsehoods, the lack of a shared information space undermines public debate. Facebook employees warned that “[t]hese ads are often so micro-targeted that the conversations on our platforms are much more siloed than on other platforms.”\(^\text{16}\) Information regulators in the United Kingdom and Spain, as well as members of the U.S. Congress, have similarly urged that platforms pause in the distribution of campaign ads. Twitter CEO Jack Dorsey announced that the company will no longer sell political ads that reference elections, candidates, parties, legislation and regulations, elected or appointed officials, or judicial decisions.\(^\text{17}\) Google has restricted micro-targeting in political ads.\(^\text{18}\) A recent survey by the Knight

---


\(^\text{16}\) Read The Letter Facebook Employees Sent to Mark Zuckerberg About Political Ads, supra note 13.


\(^\text{18}\) See Spencer, supra note 14.
Foundation and Gallup found that more than 70 percent of Americans oppose the use of personal data for microtargeting purposes by political campaigns.\textsuperscript{19}

In addition to issues with fact-checking and micro-targeting, current real-time ad labelling and after-action public libraries provide varying and inadequate information to potential voters, depriving them of the ability to know who is sponsoring ads. In the absence of industry-wide standards, platform practices differ from each other in terms of what kinds of ads they deem political. In addition, the data in the after-action databases is not robust, scoping is incorrect, identification is insufficient, metrics are fuzzy, advertisement data is unverifiable, and targeting information is lacking.\textsuperscript{20}

Journalism professor Jonathan Albright found it easy to use a false identity when buying advertisements, and Mozilla researchers found bugs and technical issues in the Facebook ad library.\textsuperscript{21} Google’s database functions better, but it does not include ads about topics, only candidates.\textsuperscript{22} Moreover, even when the information in the databases is updated and correct, it can still fail to reveal the parties funding the ads.

C. Paid Influencers and Networks of Amplifiers Have Become Critical in 2020

These networks of so-called “organic” (not ad-based) activity can flood the information zone with disinformation, manipulating algorithmic recommendations to fill trending lists and search engines. Autocratic governments have long flooded the information ecosystem to distract from inconvenient news and to deceive the public about critical or independent


\textsuperscript{20} See generally Paddy Leerssen et al., \textit{Platform Ad Archives: Promises and Pitfalls}, 8 INTERNET POL. REV. 1 (2019).


views. But this tactic formerly required considerable resources, and often ran into the roadblock of skeptical newspapers and broadcast news editors. Now, the disinformation campaign toolkit is available off-the-shelf from commercial vendors, widespread and affordable for enterprising political outfits or nation-states looking to maximize the budget lines earmarked for information warfare.

The NATO Strategic Communications Center of Excellence confirmed that it remains shockingly easy—and shockingly cheap—to purchase comments, likes, views, and followers from third parties operating on the major platforms. Private “black PR” firms increasingly offer their services to run online influence operations by using paid trolls operating fake accounts. Nathaniel Gleicher, Facebook’s head of cybersecurity policy, labeling “the professionalization of deception” as a growing threat.

Search engines are also manipulated by these actors using a variety of tactics identified by researchers, including “keyword stuffing,” or adding popular keywords to unrelated websites to promote content in search-engine rankings; “link bombs,” or increasing the number of other sites that link to the page; “mutual admiration societies,” or groups of websites with links designed to appear as legitimate citations that instead point to each other; and “data voids” that create news around an unused search term (e.g., “crisis actor” or “caravan”) and then post content with disinformation that is found by users searching for the new term.

Bots, trolls, and networks of true believers can work in coordinated fashion to increase the number of times an individual sees disinformation from different sources, crafting a sealed information environment. With such an enormous capacity to magnify a given message, repetition becomes reality. Indeed, “[t]he volume and recency of disinformation matter,” according to a Hewlett Foundation review, and “people are more likely to be affected by

inaccurate information if they see more and more recent messages reporting facts, irrespective of whether they are true.”

Amplifiers and their networks cause algorithms to sense engagement and further amplify the content they push, to the point where it emerges as a newsworthy or trending topic. These algorithms prioritize content for newsfeeds and recommendations. Search results are optimized for user “engagement” (measured by the number of comments, shares, likes, etc.) to attract and keep users’ attention so that they will stay online to be served more ads. These networks work across platforms. According to the Senate Intelligence Committee, “achieving the ‘viral’ spread of YouTube videos generally entails capitalizing on the reach and magnitude of Facebook and Twitter networks to spread links to the video hosted on YouTube.”

D. Secret Groups, Encrypted Messaging, and Fringe Sites Linking to the Main Platforms

In addition to promoting misleading news, micro-targeting users with personalized persuasion, and flooding the news zone, disinformation campaigns manipulate users by creating and infiltrating accounts, pages, and groups, pretending to represent collections of Americans with a common interest. For example, the Internet Research Agency (IRA) in Russia created a fake “Blacktivist” page that garnered 11.2 million engagements over the course of the IRA’s campaign. In general, during the 2016 elections, more than 62,000 users committed to attend 129 events organized by Russian trolls,29 including through Russian-created Facebook pages such as Heart of Texas and United Muslims of America, which had over 300,000 followers.30

But as disinformation moves to Facebook’s groups (the private version of pages) and encrypted messaging—which receives limited moderation and is not accessible to the public—even more users are susceptible to what researcher Jonathan Albright calls “shadow organizing” (when bad actors seed disinformation) without detection.31 Shadow organizing can happen across

27 Tucker et al., supra note 25, at 40.
multiple platforms—starting in fringe sites with more lenient rules such as Gab or 4Chan, and spreading to private Facebook groups and then beyond. Nina Jankowicz warns that private groups, along with the fringe sites that link to the mainstream platforms, are “where unsavory narratives ferment and are spread, often with directions about how to achieve maximum impact.”

The danger of these secluded, online propagandizing and recruitment grounds are growing increasingly apparent. For example, homegrown militia movements that traffic in conspiracy theories and refuse to recognize the authority of the federal government are organizing among members of police departments through private Facebook groups. Facebook groups for militia organizations like the Three Percenters and the Oath Keepers (which believe that the federal government plans to take away Americans’ guns, install martial law and set up concentration camps to kill dissenters), along with Neo-Confederate, Islamophobic, and white supremacist groups, count hundreds of active and former police officers among their ranks.

E. Loopholes Created by Different Platform Rules Are Applied Inconsistently and Without Transparency, Frustrating Accountability

When platforms say they do not want to police speech, they are disregarding a core part of their business. The Lawyers’ Committee for Civil Rights Under Law wrote Mark Zuckerberg that “Facebook constantly regulates speech on its platform with curation algorithms that decide which content gets amplified and which gets buried. You have decided it is acceptable to regulate speech to increase user engagement.”

Disinformation disproportionately weaponizes animosity against immigrants, Muslims, Jews, women, and African Americans. Around the world, coordinated online hate speech against racial and ethnic minorities has led to violence. Rumors about Muslims circulating on WhatsApp have resulted in hangings in India. In March 2018, the chairman of the U.N.

---


Independent International Fact-Finding Mission on Myanmar said social media companies had played a “determining role” in the violence in the country, having “substantively contributed to the level of acrimony and dissension and conflict.” These comments were echoed a year later by the U.N. Special Rapporteur, who warned that “[p]ublic institutions linked to [Myanmar’s] military, its supporters, extremist religious groups and members of the government continue to proliferate hate speech and misinformation on Facebook.”

According to the Philanthropy for Active Civic Engagement, when bad actors use various harassment techniques to “distort or drown out disfavored speech,” they disproportionally target, “journalists, women, and ethnic or racial minorities.” Since August 2019, at least three mass shooters announced their plans on a fringe website and then spread their ideology on the larger platforms.

Platforms have adopted new rules and hired tens of thousands of staff and contractors to limit hateful content, but the application and enforcement of these rules appear to be inconsistent. Leading U.S. civil rights and human rights organizations have accused Facebook of “reckless disregard for civil rights.” The Anti-Defamation League also points out inconsistency and lack

---


of transparency in enforcement as major problems." Freedom House warns that social media have "provided an extremely useful and inexpensive platform for malign influence operations by foreign and domestic actors alike." Indeed, the anti-immigration page VDare and the white supremacy newsletter American Free Press are still available. Richard Spencer remains on Twitter. Alt-right influencers and content are still widely available on YouTube, including white nationalist activist Martin Sellner, who had documented contact with the perpetrator of the mass shooting in New Zealand in March 2019. The enforcement mechanisms and moderation schemes are clearly working at a deficit, if they work at all.

The major platforms also have rules against what Facebook calls “coordinated inauthentic behavior." However, they appear to enforce these rules more consistently against foreign state operations than domestic individuals or groups. When, for example, BuzzFeed and independent researchers identified two networks of pro-Trump Facebook pages that disseminated false or misleading information in a coordinated manner, the company responded that such networks did “not violate its policy against coordinated inauthentic behavior." And the partisan and unreliable Daily

44 See Julia Carrie Wong, White Nationalists are Openly Operating on Facebook. The Company Won’t Act, GUARDIAN (Nov. 21, 2019), https://www.theguardian.com/technology/2019/nov/21/facebook-white-nationalists-ban-vdare-red-ice [https://perma.cc/MBS5-7X7G].
Wire, which garners more engagement for its content than any other significant publisher on Facebook, was found to utilize a coordinated promotion operation, yet it remains online.49

Fundamentally, the problem remains one of transparency; it is difficult to hold platforms accountable for the application of their rules, since their enforcement actions are not saved or made auditable, nor is platform traffic. For example, determining details about the Russian influence campaign in the 2016 election required the concerted efforts of both the Senate Intelligence Committee50 and Special Counsel Robert Mueller.51 In the case of airline crashes, government officials on the National Transportation Safety Board are able to collect the “black box” flight data recorder to find out what happened and help recommend updated safety regulations to the Federal Aviation Agency,52 but such post-fact analysis cannot be conducted on the platforms.

III. New Digital Democracy Architecture

Understanding how central manipulation is to the spread of disinformation allows us to craft solutions that focus on updating tried and true protections that empower users in the offline world rather than relying on top-down government control of speech or the passive hope that Silicon Valley will self-innovate its way to a satisfactory resolution.

A new architecture would ensure that companies’ policies are consistent and enforced in a manner that is clear and responsive to the public. Additionally, imposing similar obligations on similar companies would protect them from accusations of taking political sides when they take action. But this new architecture should be flexible and content- and technology-neutral without sacrificing regulatory protection or realistic enforcement options.

The new architecture would update offline laws that safeguard consumers and elections, as well as civil rights and privacy, for the online information ecosystem. It would create a fund for independent journalism,

creating the equivalent of the Public Broadcasting Service (PBS) for the Internet. It would also strengthen the old “self-regulatory” approach to Internet regulation with an industry-civil society code of conduct—focused on practices, not content—backed up by monitoring and enabled by data sharing, with a regulatory and civil enforcement backstop.

A. Update Offline Protections for an Online World

- **Use “light patterns” to empower users.** Designing user interfaces can help inform and empower users with better labeling of news, ads, altered video and audio, accounts, coordination, and even algorithmic recommendations. Platforms should provide users with the ability to customize algorithmic recommendations and track content complaints easily.

- **Restore the campaign finance bargain.** The Honest Ads Act (that would impose broadcast disclosure rules on platform ads) has bipartisan support and should set the floor for disclosure. In addition, platforms should verify who is actually funding the ads (rather than listing front groups) and platform fact-checking policies should be consistent and applied to politicians. Platforms should also, as recommended by a Federal Election Commission commissioner, limit micro-targeting of political ads.

- **Update civil and human rights law.** Discrimination, harassment, and privacy laws—such as public accommodation laws—should be updated for the digital age. Platforms should create and enforce rules that are consistent, transparent, and appealable for content removal and algorithmic prioritization.

- **Strengthen privacy rights.** The U.S. needs a uniform privacy law to provide users with the ability to protect their privacy and ensure that platforms are not allowing their data to be used to manipulate them. California has enacted a new privacy law inspired by the European Union’s General Data Protection Regulation. Meanwhile, federal privacy legislation is gaining momentum that might go beyond the “notice and consent” framework to take certain practices off the table (such as collection and sale of biometric, location, or health

---

56 2016 O.J. (L 119) 679.
information; information collected from microphones or cameras; or cross-device tracking), and to create new governance procedures for companies collecting personal information.\(^57\)

- **Stipulate national security information sharing.** Platforms should share information with each other and with government agencies on violent extremism as well as with the public on foreign election interference.

B. Provide Choice by Funding a PBS of the Internet Through a “Superfund” Type Fee on Digital Advertising Revenue

- **Platforms have syphoned away the ad revenue that once supported public interest local journalism.** A fund to support noncommercial public interest journalism, fact checkers, and media literacy could be created by taxing platform ad revenue—raising the cost of a business relying on data collection and the viral spread of disinformation while also supporting more “signal” in the system. A commitment from platforms to highlight and boost this content would also loosen algorithmic control over information flows. Just as support for public media in the past extended to communications infrastructure, so in the digital space, there should be noncommercial infrastructure as an alternative to provisioning by dominant “Big tech” companies.

C. Create Accountability With a Code of Conduct Enforceable Through a Combination of the Following:

- **Increase competition.** Lack of competition can undermine the health of the public square by limiting or skewing speech options. Policymakers understood this when they subjected broadcasters to ownership limits and prohibited them from cross-owning stations and print newspapers. Antitrust suits will move slowly and be tough to win under current law, whereas regulatory oversight can introduce more competition. Data portability would provide tools for users to export their network to competing platforms with the appropriate privacy safeguards in place. Interoperability would facilitate competition by enabling communication across networks. Some have suggested

implementation by requiring platforms to maintain APIs for third-party access under terms that are fair, reasonable, and non-discriminatory.58

- **Open the Data Black Box by mandating platforms to share data.** Platforms should provide (1) intellectual property- and privacy-protected after-action disclosure of how content is algorithmically curated; (2) what targeting policies are used; (3) moderation decision logs; and (4) access to traffic data so civil society watchdogs, researchers, and governments may help assess information flows.

- **Refocus code of conduct.** Platforms and civil society should develop a technology-neutral code of conduct focused on practices, not content.

- **Establish oversight and enforcement.** Independent data sharing would make it possible to verify that platforms are complying with the code. Monitoring can be done by independent third parties, a new Digital Democracy Board, or an existing agency with oversight authority. If necessary, Section 230 immunity59 could be conditioned on complying with the code.

### IV. Conclusion

Today, citizens themselves have few tools to evaluate a product’s security, privacy, transparency, or algorithms. The United States has abdicated its traditional leadership role on Internet policy while Europe is stepping into the void, and the Russian and Chinese governments are leveraging the lack of international consensus to use the Internet for political repression and control, both in their own countries and abroad. Meanwhile, smaller countries, left with few options, are forced to operate in a geopolitical arena with little international agreement or guidance. It is time to take active steps to ensure that the Internet is a tool to strengthen, not undermine, democratic values. In order to do so, we must agree on a common framework for understanding these challenges and embrace practical solutions that protect privacy and free expression while strengthening the information ecosystem.

---


SECTION 230

Ellen L. Weintraub* & Thomas H. Moore**

CITE AS: 4 GEO. L. TECH. REV. 625 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 625
II. THE TWENTY-SIX WORDS THAT CREATED THE INTERNET ............ 625
III. THE DISTANCE THAT INTERNET COMPANIES HAVE TRAVELED SINCE 1996 ............................................................... 629
IV. WHY FOCUSING ON SECTION 230 WILL NOT ADDRESS INTERNET COMPANIES’ INJURIES TO DEMOCRACY ........................................ 633
V. A PROPOSAL ........................................................................... 635
VI. CONCLUSION ....................................................................... 638

I. INTRODUCTION

Section 230 of the Communications Decency Act (CDA) has provided broad immunity to Internet service providers since 1996. Today, Section 230’s protections are a ripe target for those concerned about the unanticipated power and reach of the small number of giants that have emerged to dominate the digital landscape.

But for those specifically concerned about the adverse effects that Internet companies are having on American democracy—as we are—Section 230 reform is no silver bullet. Recognizing—and monetizing—Internet users’ property rights in their data may be a more productive approach.

II. THE TWENTY-SIX WORDS THAT CREATED THE INTERNET

A single sentence in the Communications Decency Act of 1996,1 Section 230, has become known as “The Twenty-Six Words That Created the Internet”.2

---

2 See, e.g., Jeff Kosseff, The Twenty-Six Words That Created the Internet (2019).

* Commissioner Ellen L. Weintraub has served on the U.S. Federal Election Commission since 2002 and chaired it for the third time in 2019.
** Thomas H. Moore serves as Chief of Staff to Commissioner Ellen L. Weintraub.
No provider or user of an interactive computer service shall be
treated as the publisher or speaker of any information provided
by another information content provider.

The provision was designed to protect fledgling Internet companies
from incurring liability when their millions of users posted content and when
the companies made moves to police that content. Sen. Ron Wyden (D-Or.),
who wrote Section 230 with then-Rep. Chris Cox (R-Cal.), says the concern
was that such liability “will kill the little guy, the startup, the inventor, the
person who is essential for a competitive marketplace. It will kill them in the
crib.”

Section 230 succeeded beyond all expectations. Amazon was just two
years old⁴ and still a precocious toddler in 1996, with revenue just shy of $16
million that year;⁵ it brought in twice as much as that every hour (for a total of
$70 billion) during the third quarter of 2019.⁶ Google, founded in 1998,⁷ two
years after Section 230 became law, had third-quarter 2019 revenue of $40.3
billion.⁸ Facebook, founded in 2004,⁹ had $17.65 billion in third-quarter 2019
revenue.¹⁰

---

⁹ Facebook, supra note 4.
The distinction that Section 230 drew between the “interactive computer services” it protects and the “information content providers” who create the services’ content has consistently withstood legal assault. Section 230 is now under political assault from legislators critical of online services’ election policies. There is plenty to be critical of: Internet companies microtarget political advertising, creating filter bubbles and preventing the counterspeech that First Amendment jurisprudence celebrates. Internet companies amplify political misinformation and disinformation. They fail to adequately protect against foreign interference in our elections. Their algorithms exploit the basic human compulsion to react to material that outrages. They have a serious problem with inauthentic users and bots that their whack-a-mole approach is not solving.

---

11 47 U.S.C. § 230(f)(2) (defining “interactive computer service” as “any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions”).

12 47 U.S.C. § 230(f)(3) (defining “information content provider” as “any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service”).


14 Facebook is the most prominent target, partly because of its size and its aggressive stance on the use of personal data. It also has the easiest financials to analyze, because data-driven advertising is by far its largest line of business. The analysis in this paper applies across Internet companies, but Facebook will stand in as the most useful example throughout.


But the most prominent political attacks on Section 230 whistle past these democracy-endangering problems and instead take aim at a straw man: the claim that the way Internet companies (Facebook, especially) moderate their content cuts against Republicans.\textsuperscript{19} There is little evidence that such partisan bias exists.\textsuperscript{20} But even if there were, there is no mandate within Section 230 for partisan neutrality,\textsuperscript{21} and any such government mandate would likely draw First Amendment-based objections.\textsuperscript{22}

Democrats have taken their swings at Facebook and Section 230 as well; Joe Biden told \textit{The New York Times}, “Section 230 should be revoked . . . [f]or [Facebook] and other platforms,” because they are “propagating falsehoods they know to be false.”\textsuperscript{23}

\textsuperscript{19} See, e.g., Nash Jenkins, \textit{The Mark Zuckerberg vs. Ted Cruz Showdown Was the Most Explosive Part of Today’s Facebook Testimony}, \textit{TIME} (Apr. 10, 2018) (quoting Senator Cruz: “There are a great many Americans who I would say are deeply concerned that Facebook and other tech companies are engaged in a pervasive pattern of bias and political censorship.”), https://time.com/5235461/mark-zuckerberg-facebook-ted-cruz/.[https://perma.cc/63K3-DES3].


\textsuperscript{22} This is illustrated well by the sharply limited ambit of the May 28, 2020, “Executive Order on Preventing Online Censorship,” Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (June 2, 2020). The E.O.’s most concrete legal step—by far—is its ordering of the Secretary of Commerce to file a rulemaking petition with the Federal Communications Commission to clarify aspects of Section 230. Filing an FCC petition is an action any person can undertake without any need for an executive order.

III. The Distance That Internet Companies Have Traveled Since 1996

When Section 230 was signed into law, online services were largely passive conduits for their users’ communications, much like a telephone company. This made sense. It would be absurd to throw a racketeering charge at a phone company every time a Tony Soprano picked up a phone to commit a crime. But what if the phone company started to listen to the content of all its calls, and used that information to deliver selected phone calls to some users and not others? What if instead of ringing its customers as soon as someone called, the phone company decided to deliver calls in an order that it determined? What if the phone rang all the time, not with calls from people known to those who answered, but from people the phone company predicted those customers might like to hear from? And that the topics those people talked about were carefully chosen by the phone company as ones that would cause emotional reactions, to drive up telephone use?

That is the direction the major online companies have taken since 1996, and some that failed to do so nimbly enough have failed entirely. Every “like,” every share, every click of every user is tracked and analyzed by online companies. Armed with this data, online companies deliver and present their information in an entirely different way. Fordham Law professor Olivier Sylvain writes:

> Intermediaries today do much more than passively distribute user content or facilitate user interactions. Many of them elicit and then algorithmically sort and repurpose the user content and data they collect. The most powerful services also leverage their market position to trade this information in ancillary or secondary markets. Intermediaries, moreover, design their platforms in ways that shape the form and substance of their users’ content.\(^\text{24}\)

What does this look like in practice? Facebook no longer distributes a user’s every post to that user’s every “friend,” or to everyone who has “liked” a business, as it once did. Instead, after spending years encouraging businesses large and small to ask their customers to “like” them on Facebook, the company turned around in 2012 and started showing its users’ posts only to a

subset of friends and likes. 25 “I felt slightly duped,” wrote Nick Bilton in a blog post for The New York Times in 2013:

“I’ve stayed on Facebook after its repeated privacy violations partly because I foolishly believed there was some sort of democratic approach to sharing freely with others. The company persuaded us to share under that premise and is now turning it inside out by requiring us to pay for people to see what we post.

Facebook takes a different view, saying that it is still finding the right balance for the algorithm that decides what people see in their news feeds. “The two aren’t related; we don’t have an incentive to reduce the distribution that you send to your followers so that we can show you more ads,” said Will Cathcart, product manager for Facebook’s news feed.26

It is worth noting that the method Facebook settled upon to allow its users to promote a post to more people was not by handwriting a polite note to its CEO, Mark Zuckerberg, but by paying cash money to Facebook.27

Facebook28 and Twitter29 no longer default to a straight chronological timeline of the material they present their users; they instead select and shuffle postings in a manner that their algorithms believe will resonate best with their users. In fact, Facebook actually bars users from bypassing its algorithms permanently.30 As Facebook executive Adam Mosseri told Time in 2015:

In general, chronological, I think, helps people who are worried about missing things or seeing more recent things, but if

30 See, e.g., Luckerson, supra note 27 (“Users can select a “Most Recent” tab to show posts as they appear, but the setting stubbornly switches back to Facebook’s algorithmically-driven feed after a certain amount of time.”).
everyone was on chronological all the time, people would miss a lot more important content. Our whole mission is to show people content that we think that they find meaningful. Recency is one important input into what people find meaningful, but we have found over and over again that it’s not the only one.31

It is likely not a coincidence that the online services’ fondness for their algorithms runs as deep as their ability to monetize these data-driven presentations. These algorithms have generated billions of dollars in advertising profits for Facebook and Google.32

The move to algorithmically driven user experiences represents a profound change in these interactive computer services. Speaking before the American Bar Association last year, Federal Trade Commissioner Rohit Chopra said:

By converging with and into behavioral advertisers, tech companies transformed their platforms into pay-to-play enterprises. Far from being a neutral or passive conduit, these platforms are now actively shaping and profiting from user communications. Gone are the incentives to attract a community by offering privacy, control, and other user-centric benefits. In their place are incentives to not only track user activity, but also generate high user activity by promoting clickbait over content shared organically by other users.33

31 Luckerson, supra note 27. Mosseri now serves as head of the Facebook subsidiary Instagram.


Section 230 aimed to protect the distribution of user-generated Internet content, but what should happen now that that content has taken a back seat to content calculated to generate outrage, stickiness to the platform, and profits? Courts are beginning to take notice of the distance Internet companies have traveled since 1996 but have so far resisted redefining platforms as content developers to pierce Section 230’s immunity protections. Last year, in *Force v. Facebook*, victims of a Hamas attack in Israel alleged that Hamas had posted content on Facebook that encouraged the attacks, suing Facebook on the grounds that its policies and algorithms “directed such content to the personalized newsfeeds of the individuals who harmed the plaintiffs.”

In July 2019, the Second Circuit held that Facebook’s direction of Hamas’ content could not be considered to render Facebook the “creator” or “developer” of the content:

> The term “development” in Section 230(f)(3) is undefined. However, consistent with broadly construing “publisher” under Section 230(c)(1), we have recognized that a defendant will not be considered to have developed third-party content unless the defendant directly and “materially” contributed to what made the content itself “unlawful.” This “material contribution” test, as the Ninth Circuit has described it, “draw[s] the line at the ‘crucial distinction between, on the one hand, taking actions... to ... display ... actionable content and, on the other hand, responsibility for what makes the displayed content [itself] illegal or actionable.”

The panel split on this question; dissenting Judge Robert Katzmann wrote: “When a plaintiff brings a claim that is based not on the content of the information shown but rather on the connections Facebook’s algorithms make between individuals, the CDA does not and should not bar relief.” Judge Katzmann’s opinion recognizes that the platforms have come to occupy an ill-defined middle ground in terms of their responsibility for the parameters of online debate—more than passive conduits but less than content originators. As increasingly sophisticated AI comes to control more and more of the content users see, prioritizing content from sources the user never chose over known sources and amplifying speech for its provocative rather than edifying content, Judge Katzmann’s cogent analysis may become more persuasive to other judges. For now, his dissenting opinion remains an outlier.

---

34 *Force v. Facebook*, Inc., 934 F. 3d 53, 59 (2d Cir. 2019).
35 *Id.* at 68 (quoting *Kimzey v. Yelp! Inc.*., 836 F.3d 1263, 1269 n.4 (9th Cir. 2016) and *Jones v. Dirty World Entm't Recordings LLC*, 755 F.3d 398, 413–14 (6th Cir. 2014)).
36 *Id.* at 77.
IV. WHY FOCUSING ON SECTION 230 WILL NOT ADDRESS INTERNET COMPANIES’ INJURIES TO DEMOCRACY

While we would welcome a broader adoption of Judge Katzmann's analysis, the nature of Internet companies’ injuries to democracy makes it unlikely that they would bear litigation liability for them, regardless of who was considered the “creator” or “developer” of the content. Section 230 jurisprudence, standing doctrine, and First Amendment law all raise obstacles to litigation.

Courts, including the majority in Force, construe Section 230 to require that an Internet company must materially contribute to the unlawful nature of the content before they will be held liable as a developer of the information. 37 Note that the underlying requirement is that the content, not the company’s practices, must be “unlawful.” 38 Unlike defamation, invasions of privacy, child pornography, terrorism, or copyright violations, the democracy-damaging information ecospheres Internet companies have created are not in and of themselves illegal. Americans deserve fair elections undistorted by Internet companies, but Congress has provided no statutory guarantee of that.

Thus, Internet companies’ democracy-damaging actions (exploiting humans’ vulnerability to outraging material, creating filter bubbles that exacerbate polarization, programming for virality, and microtargeting), being not in and of themselves illegal, could not give rise to a legal cause of action based on current interpretations of Section 230 immunity.

Another hurdle is the difficulty a plaintiff would have in establishing standing to sue in federal court. However the Internet companies’ injuries to democracy are defined, every U.S. citizen suffers that same injury. This would be the most generalized of generalized grievances.

Taxpayer standing doctrine bars a person from suing the federal government when an injury to her as a single U.S. taxpayer is identical to that suffered by her fellow 140.9 million U.S. taxpayers. 39 Likewise, any democracy-related injury suffered by a single U.S. citizen based on a platform’s distribution of information is identical to that suffered by every

37 Id. at 68–9.
38 Id. at 69.
other of the United States’s 330 million citizens.\textsuperscript{40} This standing hurdle would exist even if Section 230 were repealed outright.

Moreover, while Section 230 jurisprudence \textit{requires} a focus on the content of particular communications and its legality, such content-based analysis implicates the First Amendment. Regulating the raw materials of the damage (political mis- and dis-information) is inherently fraught and would raise concerns about the government setting itself up as the arbiter of what is and is not truth.\textsuperscript{41}

The First Amendment, as currently construed by the Supreme Court,\textsuperscript{42} also keeps federal campaign-finance law largely sidelined. Disclosure of the sources of political mis- and disinformation could and should be beefed up, as part of an enhanced disclosure regime for on-line political advertising.\textsuperscript{43} Beyond transparency measures, however, little can be added to the Federal Election Campaign Act\textsuperscript{44} by Congress, and little can be done by the Federal Election Commission that would have an appreciable effect while also passing muster with the Court.

But as campaign-finance lawyers, our instinct is to follow the money, and that instinct serves us well here. Let’s take the example of Facebook. Facebook makes its money selling ads.\textsuperscript{45} Facebook’s impressive ability to sell so very many ads is fueled by the vast amount of user data it possesses. This data allows Facebook to target, and microtarget, and adjust timelines, and to shape every bit of every user’s experience. But this data does not belong to

\begin{itemize}
  \item \textsuperscript{40} See United States & World Population Clock, U.S. Census Bureau, https://www.census.gov/popclock/ [https://perma.cc/L65W-F596] (last visited May 1, 2020).
  \item \textsuperscript{41} See, e.g., Susan B. Anthony List v. Driehaus, 134 S. Ct. 2334 (2014) (statute prohibiting false statements during a political campaign is an injury that creates Article III standing). The trial court, on remand, struck down the law: “We do not want the government (i.e., the Ohio Elections Commission) deciding what is political truth—for fear that the government might persecute those who criticize it. Instead, in a democracy, the voters should decide.” Susan B. Anthony List v. Ohio Elections Comm’n, 45 F. Supp. 3d 765, 769 (S.D. Ohio 2014). If the First Amendment means anything, it means this: the Constitution prohibits the government from persecuting those who criticize it. This bedrock First Amendment principle again reveals the fallacy underlying the May 28, 2020 “Executive Order on Preventing Online Censorship,” \textit{supra} note 22.
  \item \textsuperscript{42} Starting with Buckley v. Valeo, 424 U.S. 1 (1976), the Supreme Court has relentlessly narrowed the reach of campaign-finance law.
  \item \textsuperscript{43} See, e.g., Honest Ads Act, S. 1989, 115th Cong. (2017).
  \item \textsuperscript{44} 52 U.S.C. § 30101 \textit{et seq.}
  \item \textsuperscript{45} Sen. Orrin Hatch (R-Utah): “[H]ow do you sustain a business model in which users don’t pay for your service?” Facebook CEO Mark Zuckerberg: “Senator, we run ads.” \textit{Facebook, Social Media Privacy, & the Use and Abuse of Data: Joint Hearing Before the S. Comm. on the Judiciary & the S. Comm. on Commerce, Science, & Transportation}, 116th Cong. 21 (2018).
\end{itemize}
Facebook. The thousands of data points that Facebook has on every Facebook user\textsuperscript{46} belong to the users, not the company.

Facebook itself used to agree. But paralleling the company’s increased monetization of its users’ data are changes to Facebook’s terms of service that show an increasing unwillingness to acknowledge its users’ ownership of their personal data. What was users’ “full ownership of all of your User Content and any intellectual property rights or other proprietary rights associated with your User Content” in 2007\textsuperscript{47} has been reduced in 2020 to an acknowledgement only that users have ownership over the specific content they create and share on Facebook that enjoy copyright or trademark protection, like photos and videos.\textsuperscript{48}

Fortunately, Congress is not bound by Internet companies’ estimation of who owns their users’ personal data.

V. A PROPOSAL

One way to look at Internet companies’ businesses, their products, and the harm they are doing to democracy is by viewing that harm as a negative externality, that is, a cost of production that is shifted away from a company and paid by society instead.

Governments can reduce negative externalities by taxing or regulating them. Pollution is the classic example: A product may be cheap because the factory producing it pours its waste into a river, creating a cost to society if it wants clean water to drink. The producer can be assessed a tax to cover the

\textsuperscript{46} Data Policy, Facebook, https://www.facebook.com/about/privacy/ [https://perma.cc/H4G6-ZVC9].
\textsuperscript{48} By 2018, Facebook’s Terms of Service read: “You own the content you create and share on Facebook,” dropping the reference to “all” of a user’s content and the associated “intellectual property rights or other proprietary rights,” and narrowing coverage to things that a user might “create and share.” See Terms of Service, Facebook (last revised Apr. 19, 2018) https://web.archive.org/web/20180701035918/https://www.facebook.com/terms.php [https://perma.cc/ZD5V-7FF3]. The version current as of June 2020 defines user-owned content even more tightly. Gone is any affirmative statement that users own the content they create and share on Facebook, reading instead: “Some content that you share or upload, such as photos or videos, may be protected by intellectual property laws. You own the intellectual property rights (things like copyright or trademarks) in any such content that you create and share on Facebook and the other Facebook Company Products you use. Nothing in these Terms takes away the rights you have to your own content.” Terms of Service Facebook (last revised July 31, 2019), https://www.facebook.com/legal/terms [https://perma.cc/4AHK-4XZ9]. The terms are silent on what exactly Facebook considers to be “your own content.”
cleanup cost, or it can be regulated to keep it from creating the external cost in the first place, or both. Either course increases the cost of production (ideally, to the true cost of production) without a corresponding increase in demand, which will decrease the amount of harm produced.

Taxing Internet companies for their democracy-harming negative externalities, or banning them by regulation, is tricky business because of the First Amendment issues involved. But in this case, the negative externalities at play can also be addressed by raising the cost of inputs. If a paper mill were able to set disruptively low prices for its products because it was dumping waste into a river and it was stealing and pulping someone else’s trees, one way to reduce the harm would be to require them to pay for the trees. Internet companies’ cost of production, at the moment, includes the free use of their users’ personal data. Making them pay a market price for that input is a content-neutral way of encouraging them to make different business decisions and do less harm to democracy.

With that in mind, here’s a plan. Legislation could:

1. Recognize the property rights that users of Internet services have in their personal data.
2. Formally assign those rights to the users.
3. Require Internet companies seeking to make use of their users’ personal data to pay their users for the privilege.
4. Forbid that payment from being waived.
5. Set a fee, perhaps five percent of the payment to users—a “Democracy Dividend” of sorts—to establish a public fund aimed at enhancing democracy through such means as public campaign financing or public media.
6. Require all unclaimed payments (including those not claimed by bots and other inauthentic users) to go to the public fund.

---


This plan has the additional, perhaps even more significant, salutary effect of giving consumers control over the commercial use of their personal data.

Let’s put some rough numbers to this. What is the value of a single user’s data? Let’s use Facebook as the example again, as its financials are the most readily dissected. Loup Ventures crunched Facebook’s 2017 numbers in 2018 and found that Facebook generated in the neighborhood of $82.21 in ad revenue for every one of its active U.S. users. To get to the value of the data, Loup factored in half of Facebook’s revenue going to other costs, and it is paying a 28 percent tax rate (though Facebook works hard to pay less),\(^{51}\) for a net profit of $29.60 per active user. This is the value of a U.S. user’s data to Facebook.

Loup assumed that Facebook would pay out 70 percent of that net profit to users for their data (an assumption we also adopt) for a total of $20.72 per each active U.S. user per year.\(^{52}\)

We are not the first to propose allowing users to charge companies for their personal data.\(^{53}\) The usual criticism of such proposals is that 20 bucks a year is not enough for anyone to care much about,\(^{54}\) which may well be correct. But this plan’s main point is not Facebook’s $20 payment to individual users; it is the behavior change that could come from requiring Facebook to pay a fair price for its data inputs. Given 221 million active U.S. Facebook users,\(^{55}\) Facebook’s payments would total $4.6 billion a year for the privilege of monetizing the personal data belonging to its users. Google, Amazon, Adobe, and Verizon would likely also be among those paying their users for their personal data.


\(^{52}\) Doug Clinton, Your Data Is Worth Less Than You Think, LOUP VENTURES (Apr. 9, 2018), https://loupventures.com/your-data-is-worth-less-than-you-think/ [https://perma.cc/CL2H-2CBU]. The $29.60 valuation will come as a surprise to the forty-four percent of those surveyed, who estimated that their Facebook data was worth more than $500 per year. Id.

\(^{53}\) Such proposals even predate the Internet’s emergence as a major influence. A widely cited 1996 article by N.Y.U. Prof. Kenneth C. Laudon, for example, proposed “National Information Markets (NIMs) in which information about individuals is bought and sold at a market clearing price” to address direct-mail marketers using personal data. Kenneth C. Laudon, Markets and Privacy, 39 COMM. ACM, 92, 99 (Sept. 1996), https://dl.acm.org/citation.cfm?id=234476 [https://perma.cc/95N9-8DZ7].

\(^{54}\) See, e.g., Lee Schafer, How Much Are Your Online Data Really Worth?, PHYS.ORG (Apr. 12, 2018) https://phys.org/news/2018-04-online-worth.html [https://perma.cc/8NGB-5KMB] (noting that a $20 bill likely won’t be enough to get anyone excited about the prospect of trading away personal data to a company like Facebook.”).

\(^{55}\) Facebook’s Revenue and Net Income From 2007 to 2019, supra note 31.
The scope of the entities covered could be based on, for example, the California Consumer Privacy Act, which covers any business that does business in California and either has annual gross revenues over $25 million; buys or sells the personal information of 50,000 or more consumers or households; or earns more than half its annual revenue from selling consumers’ personal information.\footnote{Cal. Civ. Code § 1798.140.} By adopting a reasonable threshold, this kind of measure would have the additional benefit of encouraging competition in a digital world now dominated by a few giants. Only established entities would have to pay. Start-ups would be allowed some breathing room.

A public-fund fee set at five percent would generate an additional $229 million per year from Facebook payments alone. This “Democracy Dividend” could be used to strengthen democracy through the funding of civic education and digital literacy training, non-partisan fact-checking entities, public media (much as Britain’s TV licenses fund the BBC),\footnote{Licence fee and funding, BBC, https://www.bbc.com/aboutthebbc/governance/licencefee. [https://perma.cc/Z9ZV-RVEH?]} and/or public campaign financing. Payments unclaimed by fake and bot accounts would fall into this bucket as well, potentially substantially augmenting the fund and providing internet companies with a concrete incentive to reduce their ranks of inauthentic users.

Facebook’s sticking to its current plan to aggressively micro-target its users, as it appears set on doing,\footnote{See Rob Leathern, \textit{Expanded Transparency and More Controls for Political Ads}, FACEBOOK (Jan. 9, 2020), https://about.fb.com/news/2020/01/political-ads/ [https://perma.cc/H8TR-EYCC]; see also Ellen L. Weintraub (@EllenLWeintraub) TWITTER (Jan. 9, 2020, 9:20 AM) https://twitter.com/EllenLWeintraub/status/1215277203249860608 [https://perma.cc/Z28U-72EC].} would cost it $4.8 billion a year. Perhaps it stays the course and pays the price. Or perhaps it targets less advertising using personal data and can escape the charge for using that data.

VI. CONCLUSION

The role of Internet platforms has dramatically evolved from Section 230’s original conception of platforms as largely passive pass-throughs on the information highway. Today, they play an active part in shaping the outrage-inducing, fact-optional narratives that exacerbate our differences and make consensus-driven democratic decision-making harder to achieve. The platforms reap vast profits from this intervention by appropriating the value of their users’ personal data to drive the algorithms that manipulate their user experience. Restoring users’ ownership of their own data and mandating
compensation for its use will help to re-balance incentives and impose costs appropriately.

Importantly, the platforms would be paying for their conduct, in extracting and manipulating users’ personal data, not for their role in creating, disseminating, or amplifying any particular content, no matter how misleading or harmful. This content-neutral approach avoids constitutional difficulties, vindicates privacy rights, and does not require litigation to implement. It may not make the Internet a conflict-free zone, but could perhaps reduce the incentives that make it an incessant outrage machine.
TAILORING ELECTION REGULATION: THE PLATFORM IS THE FRAME

Julie E. Cohen*

CITE AS: 4 GEO. L. TECH. REV. 641 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ................................................................. 641
II. SQUARE PEGS, ROUND HOLES, AND FALSE IMPERATIVES: FRAMES FROM THE FIRST AMENDMENT’S PAST ............................................. 642
   A. From Distribution Bottlenecks to Microtargeting at Scale ............. 643
   B. From Autonomy to Automaticity ............................................ 649
   C. From “Neutral Tools” to Amplified Flows ................................... 653
III. FROM INFORMATION MARKETPLACES TO INFORMATION PLATFORMS: FRAMING THE FIRST AMENDMENT’S FUTURE ........................................ 656
IV. CONCLUSION ........................................................................ 662

I. INTRODUCTION

According to conventional wisdom, legislative efforts to limit platform-based electoral manipulation—including especially laws that go beyond simply mandating additional disclosure about advertising expenditures—are most likely doomed to swift judicial invalidation for two reasons. First, although one might wonder whether the data-driven, algorithmic activities that enable and invite such manipulation ought to count as protected speech at all, the Court’s emerging jurisprudence about the baseline coverage of constitutional protection for speech seems poised to sweep many such information processing activities within the First Amendment’s ambit.¹ Second, assuming First Amendment coverage, the level

* Mark Claster Mamolen Professor of Law and Technology, Georgetown Law. Thanks to Chris Conrad for outstanding research assistance and to Jack Balkin, Ryan Calo, Erin Carroll, Danielle Citron, Jennifer Daskal, Laura Donohue, Alexandra Givens, Ellen Goodman, Margot Kaminski, Kyle Langvardt, Amanda Levendowski, Paul Ohm, David Pozen, Neil Richards, Jessica Silbey, Gigi Sohn, and Rebecca Tushnet for their helpful comments.

of scrutiny likely to be triggered by regulation of such activities will be strict. In this Essay, I bracket questions about baseline coverage and focus on the prediction of inevitable fatality.

Legislation aimed at electoral manipulation rightly confronts serious concerns about censorship and chilling effects, but the ways that both legislators and courts approach such legislation will also be powerfully influenced by framing choices that inform assessment of whether challenged legislation is responsive to claimed harms and appropriately tailored to the interests it assertedly serves. In Part I of this Essay, I identify three frames conventionally employed in evaluating the design of speech regulation—the distribution bottleneck, the rational listener, and the intentional facilitator—and explain why each is ill-suited to the platform-based information environment, which presents different incentives and failure modes. In their place, I offer the platform itself as a new frame. Part III defines the frame more precisely, identifies the harms and interests it brings into focus, and offers some preliminary thoughts on the kinds of legislation it might permit.

II. SQUARE PEGS, ROUND HOLES, AND FALSE IMPERATIVES:
FRAMES FROM THE FIRST AMENDMENT’S PAST

In any society that uses language to communicate complex ideas, frames and framing effects are inevitable. But ill-fitting frames can engender destructive feedback loops, and that is especially true where the ongoing conversation about governance within constitutionally permissible bounds is concerned. In the course of that conversation, legislators draft to the specifications of the frames they predict courts will employ without considering whether their handiwork will address the problems they want to solve. Courts then reject such efforts when the posited relationship between proposed remedies and asserted harms does not seem to make sense. Meanwhile, legislation drafted in ways more likely to be effective often dies in committee, and predictions about what courts might do with such legislation remain untested.

The three frames described below play different roles in the contemporary First Amendment landscape, but each encourages legal actors to evaluate claims about asserted dysfunctions in (real) speech environments in ways informed by certain baseline assumptions about how such

IMS Health: Details, Detailing, and the Death of Privacy, 36 VT. L. REV. 855, 859–61 (2012); but see Neil Richards, Why Data Privacy Law Is (Mostly) Constitutional, 56 WM. & MARY L. REV. 1501 (2015). I disagree that the First Amendment does or should apply to information processing activities regardless of their nature and context, but that is a subject for a different occasion. See generally Frederick Schauer, The Politics and Incentives of First Amendment Coverage, 56 WM. & MARY L. REV. 1613 (2015).
environments work. Each operates by reference to a familiar ideal of competition and contest according to which, as Justice Holmes put it, “the best test of truth is the power of the thought to get itself accepted in the competition of the market.” More importantly for purposes of this Essay, each focuses attention on particular kinds of market failure and suggests correspondingly particular criteria for market success, and each assumes certain structural preconditions within which market dynamics unfold.

A. From Distribution Bottlenecks to Microtargeting at Scale

One frame conventionally employed in evaluating speech regulation is the idea of the distribution bottleneck. A distribution bottleneck confers market power on whoever controls it, but the frame of the distribution bottleneck is not concerned with market power in the abstract. It is an artifact of mid-twentieth-century media regulation and litigation, and so it is also an artifact of the principal risk to free expression that mid-twentieth-century media technologies were thought to create: centralized, practically and technically unavoidable control of access to communication channels resulting in preemptive censorship. In the late twentieth century, as media technologies evolved and the power of media ownership began to manifest in ways that did not align with the frame, interested actors mobilized the frame to mount successful campaigns for deregulation. In the Internet era, platforms exercise power in ways that do not appear within the frame at all.

The distribution bottleneck frame originated in disputes about the constitutionality of regulations designed to limit the power of mass media owners. So, for example, because the then-usable broadcast spectrum imposed a natural bottleneck effect, the Federal Communications Communication imposed an access mandate—the fairness doctrine—on broadcast licensees to ensure that those wishing to express opposing or minority viewpoints had opportunities to respond to certain kinds of statements. In practice, the doctrine proved controversial, opening new vistas for gamesmanship and

---


threatening to embroil regulators directly in content disputes, and the FCC ultimately withdrew it.\(^5\) Even after the fairness doctrine’s demise, however, the distribution bottleneck frame survived. As technologies for cable and satellite distribution introduced multiple alternative channels for reaching viewers, courts began to rely on the frame to strike down new types of media regulation, reasoning that market pressures would ensure adequate alternative avenues of communicative opportunity. In response to such decisions, regulators gradually learned to color within the boundaries that the bottleneck frame imposed.\(^6\)

The distribution bottleneck frame informs the modern landscape of anti-electioneering jurisprudence in two ways. First, the Court’s evolving stance on the constitutionality of limits on campaign contributions reflects an analogous understanding of the relationship between scale and control. The earliest decisions upholding contribution limits painted large contributions as inevitably leading to corruption of the democratic process because they engendered a “pay-to-play” norm.\(^7\) According to that way of thinking, money deployed at scale functions in the manner of a bottleneck limiting access to political influence. It crowds out disfavored inputs to political decision-making, and it does so explicitly in the service of particular outcomes. Later decisions reversing course on the constitutionality of limits on independent expenditures characterized democratic politics as inherently transactional, observing that “[a]ll speakers, including individuals and the media, use money

---

\(^{5}\) See Inquiry into Alternatives to the General Fairness Obligations of Broadcast Licensees, 102 F.C.C.2d 143, 147–48, 246 (1985) (concluding “that the fairness doctrine, as a matter of policy, disserves the public interest...”); Syracuse Peace Council v. FCC, 867 F.2d 654, 669 (D.C. Cir. 1989) (upholding the FCC’s order abolishing the fairness doctrine).

\(^{6}\) See Reno v. ACLU, 521 U.S. 844, 853 (1997) (observing that “[n]o single organization controls any membership in the Web, nor is there any single centralized point from which individual Web sites or services can be blocked from the Web.”); Denver Area Educ. Telecomms. Consortium v. FCC, 518 U.S. 727, 776–78 (1996) (Souter, J., concurring) (discussing the difficulty of ruling on the constitutionality of Congress’s permissive grant of authority to cable operators to regulate the content of leased independent programmers, given the “technological and regulatory flux” of industries in which its component “individual entities [can] act as bottlenecks to the free flow of information.”); Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 663 (1994) (upholding Sections 4 and 5 of the Cable Television Consumer Protection and Competition Act of 1992 while observing that the “First Amendment’s command that government not impede the freedom of speech does not disable the government from taking steps to ensure that private interests not restrict, through physical control of a critical pathway of communication, the free flow of information and ideas”); Ellen P. Goodman, Media Policy and Free Speech: The First Amendment at War With Itself, 35 Hofstra L. Rev. 1211, 1226–27 (2007).

amassed from the economic marketplace to fund their speech.”

From that perspective, the bottleneck effect disappears. Campaign finance arrangements merely express the “power of the thought to get itself accepted in the competition of the market” and thereby mirror the Darwinian struggle for supremacy that Justice Holmes envisioned.

Second, the bottleneck frame helps to explain how modern regimes of disclosure-based election regulation have chosen to handle the problem of anonymous speech. After *McIntyre v. Ohio Election Commission*, in which the Court invalidated a state prohibition on anonymous leafletting, many states amended their laws. They did not eliminate speaker identification requirements, but rather crafted narrow exceptions permitting anonymous election-related speech by relatively small-scale speakers. Through the lens of strict scrutiny—or even that of “exacting scrutiny” as articulated in the Court’s later election regulation cases—that resolution is difficult to understand. The concerns articulated by the *McIntyre* majority about chilling effects and failure to differentiate the messenger from the message apply equally to large, popular, and well-resourced actors. If statutory shelters for electioneering speech by small-scale speakers make sense at all, that can only be because (largely implicit) preconceptions about the necessary correlation between size and electoral influence render the state interest in disclosure about small-scale interventions much less compelling.

Platforms that combine networked economies of scale with capabilities for data-driven, algorithmic microtargeting and socially networked, cascading flows of information restructure the relationships between money, scale, and the possibility of improper influence in ways that defy earlier assumptions. Begin with bottleneck control. Arguably, today’s dominant advertiser-funded platforms qualify as distribution bottlenecks for content, but in other respects the analogy to the types of control enjoyed by mid-twentieth-century

---


13 *McIntyre*, 514 U.S. at 357.
television and radio networks is difficult to sustain. Platforms like Google, Facebook, and Twitter have thrived precisely because they enable certain forms of access to the so-called long tail—i.e., content of interest to only a small number of readers. As long as they are willing to use the dominant platforms’ services, thereby foregoing direct access to both user data and information about algorithm design and training, would-be speakers of all sizes and persuasions can buy targeted advertising at a relatively low cost. Platforms might exercise preemptive control of the content of such ads, but for the most part they do not, as that would be much less profitable. Their interest lies simply in extracting surplus from whatever types of messaging elicit responses (positive or negative, but in any event data-generating) from their customers.14

Even so, the combination of platforms’ own economic self-interest and the narrower interests of those who purchase and compete for digital advertising reshapes the universe of information available to users. Platforms win when they can promise the most comprehensive and accurate methods of targeting content based on predicted interest and the largest pool of potential viewers of that content; advertisers win when they achieve clickthrough, and content providers win when they can promise advertisers higher pageviews via either targeting or social sharing of their content. For platforms, competition for eyeballs both incentivizes and rewards interface design that keeps users on the platform and tracks them carefully and comprehensively as they browse, click, like, hate, comment on, and share items with one another.15 For advertisers and content providers, competition for eyeballs both incentivizes and rewards content design for maximal “engagement” as defined by those activities.16 The resulting effects have been termed “filter bubbles,”


16 See Franklin Foer, World Without Mind: The Existential Threat of Big Tech
but that term is to some extent misleading. Platform users do not experience or self-select into impermeable bubbles but rather sort themselves into opposing tribes. They respond most readily and predictably to content that reinforces their tribal inclinations—especially content that triggers outrage or affords opportunities to signal affiliation—and they search for content using syntax that prompts algorithms to serve up tribally validating results.\footnote{See Alice E. Marwick, Why Do People Share Fake News? A Sociotechnical Model of Media Effects, 2 GEO. L. TECH. REV. 474 (2018); Francesca Tripodi, Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices, DATA & SOC’Y (May 16, 2018), https://datasociety.net/library/searching-for-alternative-facts/ [https://perma.cc/2QKK-3NYK]. The “filter bubble” terminology originated with Eli Pariser, The Filter Bubble: What the Internet Is Hiding From You (2011).}

Market dominance plays a role in this story—platforms win most decisively when they can promise the largest pools of potential viewers for any and all content—but disrupting the dominance of any particular platform would not cure the dysfunctions that more widely distributed capabilities for personalization at scale and optimization for engagement now create. In a networked media ecosystem designed for content targeting, optimization for engagement, and amplification of social flows, polarized and polarizing content spreads rapidly from one platform to another and between online and traditional media, gaining in volume as it travels.\footnote{On the interconnectedness of online and traditional media, see generally Ulrike Klinger & Jakob Svensson, The Emergence of Network Media Logic in Political Communication: A Theoretical Approach, 17 NEW MEDIA & SOC’Y 1241 (2015). See also Erin C. Carroll, News as Surveillance, 59 WASHBURN L. J. (forthcoming 2020).}

Under such conditions, the implicit presumption about the relative inefficacy of small-scale interventions also no longer holds. Because information flows within platform-based, massively intermediated environments are data-driven and social, provocations from the margins can be designed to trigger patterns of rapid, cascading spread. Such provocations exploit properties of human behavior—most notably, fear of missing out on what everyone else already knows; properties of social networks—particularly
their hub-and-spoke organization, which permits rapid spread via well-connected nodes; and properties of organizational behavior—especially traditional media outlets’ eagerness to chase and report on topics trending online.19 Multiple teams of researchers studying election manipulation have mapped the resulting patterns, tracing the paths followed by extreme and inflammatory content as it migrates from the periphery to the center of public consciousness.20 Some such interventions originate with well-resourced state actors and powerful domestic political blocs, but others have been true bottom-up efforts.21

Because the distribution bottleneck frame originated in a world characterized by hierarchical control of content prepared for distribution to mass audiences, it has little of direct significance to say about either the operation or the distinctive dysfunctions of platform-based, massively intermediated information environments. It comprehends neither the sorts of personalized microtargeting that platform-based information infrastructures enable nor the ways that optimization for data-driven surplus extraction and competition for eyeballs incentivize self-sorting into political tribes hardened in their mutual contempt for one another. Yet it constitutes an imagined world in which the very possibility that third parties might hijack and weaponize socially networked flows is already foreclosed. Legislators attempting to craft new anti-electioneering laws for the platform era and courts reviewing such efforts should recognize that the distribution bottleneck frame has no place in either exercise.


B. From Autonomy to Automaticity

A second frame conventionally employed in designing and evaluating speech regulation is the idea of the rational listener. The rational listener tests ideas for their persuasiveness and vets factual propositions for their truthfulness but has no interest in efforts to impose general, ex ante restrictions on the flow of low-quality ideas and propositions. The rational listener is autonomous and perspicacious and therefore (largely) self-reliant, capable of separating fact from falsehood and reason from self-interested conniving and demagoguery. Legislative design for the rational listener accordingly emphasizes transparency and informed choice, and courts have tended to regard such approaches as acceptable ways of advancing state interests precisely because they leave room for rational listeners to make their own decisions.

The idea of the rational listener has deep roots in the Anglo-American political tradition and more direct and immediate roots in the American system of political economy. Its first judicial articulation emerged in early twentieth-century cases involving restrictions on political liberty. The contemporary rational listener frame, however, is also and importantly an artifact of mid-twentieth-century litigation over economic and consumer protection regulation. Thus, it is also an artifact of the particular risks that mid-twentieth-century consumer markets were thought to create—risks involving the emergence of more complex consumer products and services that consumers themselves could not easily evaluate. Regulators responded to those developments by prohibiting certain kinds of deception and requiring certain kinds of disclosure. The rational listener frame, however, dictated that ultimate decision-making authority should remain with the individual to the greatest extent practicable. Courts therefore struck down laws regulating advertising, labeling, and similar matters that seemed to be attempts to superimpose government judgments about the ultimate desirability of the covered products and services.

clear preference for laws and regulations that focused simply on injecting more or different kinds of information into the marketplace and into public discourse.

Within the modern landscape of anti-electioneering jurisprudence, the rational listener frame is especially prominent in disputes about laws mandating disclosure of campaign contributions and advertising expenditures. In an era when the anti-corruption rationale for upholding spending limits no longer holds sway, federal election regulation depends ever more heavily on such provisions. Litigation over their constitutionality has given rise to the idea of an “informational interest” that is sufficiently important to warrant (slightly) relaxed scrutiny—“exacting” rather than “strict”—and also to override anonymity interests in certain circumstances. Over time, the Court’s opinions elaborating the informational interest have leaned heavily on the frame of the rational listener. In particular, within the more recently developed conception of purchased access as consistent with a broadly transactional democratic politics, disclosure “enables the electorate to make informed decisions and give proper weight to different speakers and messages.”

Many current reform proposals for tackling platform-based disinformation double down on transparency, proposing to require disclosures about a variety of matters including ad buys, ad targeting, and automated “bot” speech. In the abstract, such proposals sound like great ideas. One might even hope for new, technologically mediated advances in electoral transparency. Networked information technologies have already facilitated widespread, easy access to data about contributions to political campaigns; now, machine learning techniques can be trained on other categories of disclosed data to map networks of influence. Rational listeners who discover that they have been consorting with bots can reevaluate their choices.


26 *Citizens United*, 558 U.S. at 371.


Optimism about the potential for greater disclosure to ameliorate the problems caused by microtargeting is misplaced, however, because platform-based, massively intermediated information environments are not designed for the rational listener. Instead, they are both systematically configured and continually reoptimized to elicit automatic, precognitive interactions with online content. As noted in Section A, the currency of the platform-based environment is user behavioral data, and that reality dictates a set of interrelated strategies for platform providers. Platform interfaces work to normalize consent to tracking; to keep users on the platform to facilitate the most comprehensive tracking; and to harvest data about user preferences and aversions using low-level stimulus-response feedback loops—e.g., “buttons” for liking and sharing content—designed for automatic, habitual engagement.29 Behind the scenes, platform algorithms work to derive behavioral and psychographic profiles based on user engagement data; to drive socially-networked flows of content; and to amplify such flows in ways that maximize advertising revenues.30 And, as noted above, these characteristics of platform-based environments engender complementary strategies for advertisers and content providers, who work to design “clickbait” and foster its widest possible circulation.

Skeptics and advocates seeking to minimize alarm about the effects of platform capabilities for voter microtargeting argue that microtargeting promises more than it delivers because it cannot change minds, but that

---


argument mistakes the purposes for which microtargeting is more commonly deployed by political operatives. Without a doubt, inducing undecided (or actively hostile) voters to vote for a particular candidate is much harder than inducing them to order of-the-moment, celebrity-endorsed sneakers or book an ostensibly discounted stay at a luxury hotel (“only three rooms left at this price!”). Using behavioral and psychometric targeting techniques to play on recipients’ fears and to activate their tribal loyalties and enmities, however, is a different—and much easier—proposition.

The result of platform design for maximal data harvesting, continual user engagement, and cascading, socially networked spread based on automatic, conditioned responses is a networked digital environment in which the rational listener’s presumptive autonomy increasingly is displaced by automaticity—by habitual, precognitive behaviors that require no conscious attention. Platform-based environments constitute what legal philosopher Mireille Hildebrandt terms the digital unconscious, a field of operation within which agency is mindless, data-driven and characterized by “ubiquitous anticipation” of user predispositions. The individual subject of the digital unconscious is not the rational listener but rather the listener who is not really listening at all. Critically, moreover, the digital unconscious is also a “a field of operation for precognitive activation and manipulation at scale.” Voter microtargeting efforts move and are designed to move on the collective level, nurturing rumor and innuendo, hardening targeted populations in their tribal responses to real and perceived differences, and frustrating the sorts of efforts toward rapprochement on which theories about republican self-government rely.

The rational listener frame, which foregrounds the autonomous individual, cannot make sense of the platform-based information environment. Regulatory initiatives based on mandated disclosure, which are oriented toward the needs and presumed competencies of the rational listener, fatally misapprehend platforms’ operative logics and scalar effects. So too with solutions based on fact-checking by third parties, whose interventions must battle upstream against an unrelenting torrent of bias reinforcement, and those based on opt-out rights, which rely on recipients themselves to recognize and disavow their own most automatic and deeply-ingrained habits and affinities. Legislators attempting to craft new anti-electioneering laws for the platform

era and courts reviewing such efforts should understand that the rational listener cannot help them.

C. From “Neutral Tools” to Amplified Flows

The third frame conventionally employed in evaluating speech regulation—one that comes into play when the legal responsibility of third-party intermediaries is at stake—is the idea of the intentional facilitator. According to this frame, a third-party intermediary should not automatically incur liability for harms caused by information circulated by others. For both legislators and courts, that prospect raises worries about censorship by proxy. The intentional facilitator frame counteracts those worries by linking liability to some type of knowing involvement with specific items or categories of clearly illegal content.

The intentionality frame powerfully infuses two very different statutes governing responsibility for online content that are widely understood as encoding opposite policy choices. One is Section 230 of the Communications Decency Act, which immunizes interactive service providers from liability for unlawful content published by users of their services unless they have played a role in its development.\(^{33}\) The other, Section 512 of the Copyright Act, creates safe harbors for information intermediaries but withholds safe harbor from intermediaries that have failed to act upon receiving knowledge of specific infringing content or that have offered services specifically designed to profit from infringing flows.\(^{34}\) Intentionality plays a central role in both regimes even though the conditions for loss of immunity differ. In particular, although Section 230’s drafters sought to limit the effect of background doctrines tying liability to mere knowledge, and Section 230’s contemporary defenders regard the Copyright Act’s notice-and-takedown regime as antithetical to Section 230’s animating spirit, both statutes link liability for facilitating the spread of harmful content to volitional involvement with specific content rather than to the underlying design of distribution mechanisms more generally.\(^{35}\)


\(^{34}\) 17 U.S.C. § 512(c)-(d) (2012).

\(^{35}\) On the intent behind Section 230, see H.R. Rep. No. 104-458, at 190 (1996) (“The conferees intend that [CDA Section 230’s] defense be construed broadly to avoid impairing the growth of online communications through a regime of vicarious liability.”); see also id. at 194; Robert Cannon, The Legislative History of Senator Exon’s Communications Decency Act: Regulating Barbarians on the Information Superhighway, 49 FED. COMM. L.J. 51, 61–70 (1996). For representative contemporary reactions to the prospect of replacing Section 230 with a notice-and-takedown regime, see DANIELLE CITRON, HATE CRIMES IN CYBERSPACE 171–89 (2014); Mike Masnick, Thanks To Copyright, We Already Know How Aggressive Content Moderation
Notably, courts evaluating disputes under both Section 230 and Section 512 have homed in on the underlying commonality, framing networked digital technologies—including platform-based, massively intermediated information environments—as neutral tools that are, and should be, exempt from more intrusive oversight. So, for example, in cases about the scope of Section 230’s immunity, courts have opined that an online dating service that failed to implement certain safety features could not be penalized because it simply “provid[ed] ‘neutral assistance’ in the form of tools and functionality available equally to bad actors and . . . intended users”; and that a roommate matching service could not be responsible for discriminatory requests posted in the spaces it provided for unstructured comments because liability for providing a “simple, generic prompt” would be inconsistent with the immunity afforded to services “that provide users neutral tools to post content online.” In cases about the scope of Section 512’s safe harbors, they have rejected interpretations that would impose liability based on general awareness of likely infringement because such interpretations would require platforms to monitor their systems for signs of illegality.

In the abstract, there are sound policy reasons for worrying about the effects of liability for tool developers. In particular, courts and commentators worry with good reason that takedown obligations could morph into an open-ended mandate to sanitize the universe of public information by removing controversial content. Concerns about giving copyright interests de facto control over technological development also are well taken. It is worth noting, though, that the quasi-religious devotion to untrammeled innovation that sometimes accompanies such concerns is both an historical anomaly and


37 Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1174–75 (9th Cir. 2008) (en banc).

38 UMG Recordings, Inc. v. Shelter Capital Partners LLC, 718 F.3d 1006, 1022–23 (9th Cir. 2013); Viacom Intern., Inc. v. YouTube, Inc., 676 F.3d 19, 35 (2d. Cir. 2012).


an anti-regulatory dog whistle. Major sectors of the modern regulatory state emerged precisely to constrain innovation’s excesses, and although the design of regulatory oversight mechanisms has engendered profound disagreements, support for such basic propositions as, say, the continued existence of agencies devoted to environmental protection and food and drug oversight is broad and durable.

The “neutral tools” characterization, however, is overly simplistic both in general and as applied to platform-based, massively intermediated information environments. Generally speaking, tools reflect the priorities of their designers and may disserve or simply overlook other priorities and needs.41 We have already seen that the platform-based, massively intermediated environment alters the universe of information in specific, non-neutral ways. Platforms’ formal agnosticism about information content belies an operational orientation that reliably infuses information flows with distinctive attributes. The relatively crude distinction between knowing involvement and the mere provision of neutral tools for accessing information elides design principles that privilege polarization, amplification, and automaticity, and those principles shape both the content and the consumption of networked, massively intermediated communication.

In the context of platform-based, massively intermediated environments, the legal system should be less concerned with intentionality as to specific pieces of content—a lens that inevitably implicates the state in choice of political preferences—and more concerned with a deliberate design orientation that privileges automatic, habitual response and reflexive amplification. As currently constituted, the platform-based, massively intermediated information environment is an arena for Darwinian struggle in which the determinant of superiority is not truth but rather bias confirmation. The First Amendment does not require legislators or judges to privilege design for automaticity and reflexive amplification, and it permits them to select a frame that makes such choices and their undeniable, empirically demonstrated effects more salient.

41 For a good general introduction to the social construction of tools and technologies, see Wiebe E. Bijker, Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change (1995). Cf. Daphne Keller, Toward a Clearer Conversation About Platform Liability, KNIGHT FIRST AMEND. INST. (April 6, 2018), https://knightcolumbia.org/content/toward-clearer-conversation-about-platform-liability [https://perma.cc/E79Q-6PDS] (“All of this makes neutrality something of a Rorschach test. It takes on different meanings depending on the values we prioritize.”).
III. FROM INFORMATION MARKETPLACES TO INFORMATION PLATFORMS: FRAMING THE FIRST AMENDMENT’S FUTURE

Understanding the ways that platform-based, massively intermediated information environments work, and the ways that such environments engender unacceptable structural conditions for public discourse, suggests a new frame to be used in designing and evaluating speech regulation: that of the platform seen for itself. The interests implicated by this frame are not simply informational, and the compelling need to protect them justifies both new types of regulatory oversight and new ways of thinking about the associated tailoring problems.

It is useful to begin with definitions. As applied to networked information intermediaries, the term “platform” is a metaphor, one that has worked both to draw attention to certain features of platform-mediated spaces and deflect attention from others. Sustained scrutiny of information platforms, however, has surfaced more information about their attributes and capabilities, making it possible to describe those attributes and capabilities in more precise ways that could inform new framework legislation. An information platform is an information intermediary that uses data-driven, algorithmic methods and standardized, modular interconnection protocols to facilitate digitally networked interactions and transactions among its users. As that general definition is intended to suggest, a platform-based environment might be designed in a variety of ways. This Essay, however, has identified the following capabilities that have become characteristic of contemporary platform-based environments: collection of highly granular data about user behaviors; design of interfaces to elicit behavioral data via automatic, conditioned responses; processing of such data to create behavioral and psychometric profiles of users and user populations; targeting of content to users and user populations based on such profiles; and algorithmically-mediated amplification of content based on user engagement.

The definition articulated above also makes clear what platforms are not: they are not publishers, nor are they public fora as that concept has conventionally been understood and elaborated within First Amendment jurisprudence and theory. Platforms are private, for-profit entities that

43 For representative examples of such arguments, see Eric Goldman, Section 230 Applies to Facebook’s Post Removals and Account Suspensions–King v. Facebook, TECHDIRT BLOG (Sept. 6, 2019), https://blog.ericgoldman.org/archives/2019/09/section-230-applies-to-facebooks-post-removals-and-account-suspensions-king-v-facebook.htm [https://perma.cc/TDU7-3HYS]; Mike Masnick, Supreme Court Signals Loud and Clear That Social Media Sites Are not Public Forums That Have to Allow All Speech, TECHDIRT (Jun.
operate as central nodes in the contemporary personal data economy. They afford their users opportunities for self-expression because self-expression generates behavioral data that can be monetized. They route content (or, more accurately, links to content published by others) using predictive algorithms that have been trained on user behavioral data, and they amplify socially networked flows in ways that elicit conditioned, automatic, and tribal responses because that is the approach that most reliably enriches their shareholders and venture investors. Seen for themselves, platforms merit neither the solicitude traditionally accorded publishers wishing to express their opinions nor the rote, unthinking application of rules traditionally applied to institutions performing public access functions.44

Platform capabilities do not simply threaten the informational interest long recognized in the Court’s election jurisprudence. They also threaten other interests that are important both instrumentally—i.e., as ways of ensuring fidelity to the informational interest—and intrinsically because they are inseparably intertwined with preservation of a system of government that is accountable both to the people it serves and to the rule of law. (To be clear, platform capabilities also implicate interests that I do not discuss here. For example, anti-vaxxer propaganda that risks undermining herd immunity jeopardizes an important interest in public health. For purposes of anti-electioneering regulation, however, the interests described below are key.)

The first interest threatened in platform-based, massively intermediated information environments is an anti-factionalism interest. As Anthony Johnstone has explained, such an interest is both latent in some strands of contemporary anti-electioneering jurisprudence and solidly grounded in an original understanding of the Constitution.45 In The Federalist No. 10, Madison cautioned explicitly and pointedly against the threat posed by factions that might first capture and then subvert the institutions of democratic government by subordinating public functions to their own

---

44 I intend no comment on whether, having opened social media accounts, government officials must manage those accounts in a manner consonant with public forum doctrine. That question is both analytically distinct from questions about the status of platforms themselves and far more amenable to straightforward doctrinal analysis. See Knight First Amend. Inst. v. Trump, 928 F.3d 226 (2d Cir. 2019); Davison v. Randall, 912 F.3d 666 (4th Cir. 2019). But see Morgan v. Bevin, No. 3:17-CV-00060-GFVT-EB, 2018 U.S. Dist. LEXIS 204657 (E.D. Ky. Dec. 3, 2018) (confusing the two questions).

narrower interests. Modern arguments for translation of the anti-factionalism interest into the domain of election regulation have focused on reviving the broader, now-disfavored, understanding of campaign finance as part of a conscious return to civic republicanism. But the anti-factionalism interest also bears on the ongoing debate about the structural properties of platform-based speech environments. The centrifugal properties of the platform-based environment—within which communications are systematically optimized to elicit, separate, and harden tribal reflexes—enable powerful factions to weaponize networked information flows in order to perpetuate their own power and advantage. At the same time, they disable the collective capacity to produce and propagate gap-bridging responses.

The second interest threatened in platform-based, massively intermediated information environments is an anti-manipulation interest. As defined by Daniel Susser, Beate Roessler, and Helen Nissenbaum, manipulation means hidden interference that deprives us of authorship over our own choices. As Susser, Roessler, and Nissenbaum argue, if a rule against manipulation is to have any concrete force, it must apply to the structure of the networked communications environment rather than just to particular, discrete communications that contain manipulative content. Manipulation in platform-based information environments is neither occasional nor accidental; it is endemic and results from capabilities that platforms systematically design, continually reoptimize, and deliberately offer up to third parties for exploitation. Properly conceived, the anti-manipulation interest encomasses the dark patterns that keep users enrolled and logged in, the stimulus-response loops designed to elicit automatic, precognitive responses and harvest the resulting data, and the mechanisms for harnessing that data to enable microtargeting based on user vulnerabilities and fears.

Finally, the emergent properties of information flows in platform-based, massively intermediated environments threaten a structural interest that warrants separate recognition. This interest, which I will call an anti-authoritarianism interest, concerns the stability and robustness of foundational democratic institutions and requires us to confront another underlying presumption of the marketplace-of-ideas metaphor that underwrites so much of First Amendment jurisprudence and theory. Implicit

46 See The Federalist No. 10 (James Madison).
47 See, e.g., Zephyr Teachout, Corruption in America: From Benjamin Franklin’s Snuff Box to Citizens United (2016); Lawrence Lessig, Republic, Lost: How Money Corrupts Congress—and a Plan to Stop It (2011); see generally Lawrence Lessig, Fidelity in Translation, 71 Tex. L. Rev. 1165 (1993).
49 See id. at 38–41.
in the marketplace-of-ideas story is an optimistic prediction about what will happen when open information systems (“more information”) and authoritarian information systems (“censorship”) collide: As Darwinian conflict kicks in, truth (and by extension democratic self-determination) will prevail over falsehood (and by extension autocracy).\(^{50}\) As Henry Farrell and Bruce Schneier show, that is too simple. Authoritarian information systems have developed sophisticated information strategies that leverage platform-based environments to undermine common knowledge about how democratic institutions function and, by extension, to destabilize the behavioral norms that lend such institutions continuing legitimacy.\(^{51}\) Such attacks, which are now well-documented, exploit platform capabilities for microtargeting, automaticity, and cascading, socially-networked information spread to stoke conspiracy theories and foster distrust—of government, of the “mainstream media,” of scientific consensus around topics such as climate change and the efficacy of vaccines, and so on.\(^{52}\) Powerful domestic factions that should have mobilized to defend these assaults on our foundational institutions instead have adopted weaponization techniques to further their own ends.\(^{53}\) As such strategies become more powerful, they produce and amplify modes of public discourse about institutional actors that are incompatible with the knowledge structure of a stable democracy.


\(^{52}\) See S. SELECT COMM. INTELLIGENCE, 116TH CONG., REP. ON RUSSIAN ACTIVE MEASURES, CAMPAIGNS, AND INTERFERENCE IN THE 2016 U.S. ELECTION (2019); Bradshaw & Howard, supra note 20; Hindman & Barash, supra note 20; see also MARGARET E. ROBERTS, CENSORED: DISTRACTION AND DIVERSION INSIDE CHINA’S GREAT FIREWALL (2018).

Each of these interests is compelling enough in its own right to warrant some degree of regulatory oversight. Their cumulative weight is considerably greater. But we have now arrived squarely at the problem of tailoring. Are there regulatory avenues that would safeguard the interests I have identified without doing violence to others that are equally important? Drafting such legislation is beyond the scope of this essay. Drawing on the analysis in Part I, however, I want to suggest two general sets of guidelines.

First, proposed legislation that is touted as targeting the dysfunctions of the platform-based, massively intermediated environment should stand or fall based on whether or not it actually does so—whether it responds to the failure modes of the platform rather than to abuses of distribution bottlenecks, to the types of remediable information complexity that frustrate the rational listener, or to the transparent venality of the intentional facilitator. Put differently, we should not expect interventions directed only toward the largest platforms, or only toward enabling individual choice about targeting, or only toward expanding DMCA-style liability or liability based on “reasonable efforts” at post hoc content removal, to accomplish much.54

For similar reasons, platform initiatives for self-governance via “content moderation” should be understood for what they are: shiny, expensive distractions designed to stem the rising tide of criticism without undercutting the core platform business model, which depends on the relative profitability of immoderation. Oversight boards, internal appeal processes, and the like appeal to the lawyerly taste for process, but their significance is more performative than real.55

By contrast, I have identified three structural features of platform-based intermediation that threaten the anti-factionalism, anti-manipulation, and anti-authoritarianism interests: predictive profiling and microtargeting based on behavioral and psychographic data; interface design to elicit automatic, precognitive responses; and algorithmic optimization to amplify patterns of cascading, socially-networked spread. Each of these features is amenable to systemic oversight, audit, and intervention, and platforms’ own actions confirm this. As platforms doggedly pursue ever more intrusive forms of behavioral and psychographic profiling, refine their interfaces to enable ever more seamless collection of user feedback, and continually tweak their algorithms to optimize both viewer engagement and networked information

---


55 For additional development of this point, see COHEN, supra note 14, at 135–36, 249–50.
spread, they also give the lie to the oft-repeated canard that their actions are intrinsically ungovernable. It may well be “impossible to do content moderation well,” but it is not impossible to imagine regulation targeted to those very different and more systemic failure modes, nor should it be beyond the pale of civil conversation among twenty-first century civil libertarians to do so.

A second set of relevant guidelines concerns the relative importance of different kinds of tailoring errors. All current versions of First Amendment scrutiny presume that the costs of mistaken instances of suppression (far) outweigh those of mistaken failures to suppress. That preference in turn rests on important assumptions about the nature and operation of the information environment—most notably, that injecting more speech into the marketplace is costly and that instances of low-value speech are readily ascertainable either by the rational listener or via intermediaries whose claims to authority the rational listener can readily assess. Those are not the properties of platform-based, massively intermediated information environments, and so the underlying presumption about error costs may warrant revisiting. As Frederick Schauer has explained, the First Amendment’s costs have always been distributed unevenly. As long as those costs did not threaten the overall


58 See generally Frederick Schauer, Harm(s) and the First Amendment, 2011 SUP. CT. REV. 81, 108–10 (2011).
stability of a system of democratic government accountable to the people and to the rule of law, however, they could be written off as the sort of collateral damage inevitable in a constitutional system designed to privilege liberty over equality and anti-subordination. Now that overall stability is on the table, however, it may be worth asking new and more probing questions about harms and costs.

Policymakers wanting to engage in a sensible discussion about tailoring and error costs, however, should remember that the project at hand entails designing effective oversight of behavioral conditioning and algorithmic amplification, and that the most effective forms of oversight will not consist of cumbersome, user-driven mechanisms for post hoc content removal. So, for example, we might begin by asking whether and under what circumstances we should agree to trade reduced scope for the viral spread of grass-roots political dissent against reduced scope for the viral spread of messaging about the need for armed insurrection in response to purported racial “replacement” or purportedly “rigged” elections. Properly understood, though, those questions are not about whether to jettison long-established principles designed to preserve breathing room for dissent. Rather, they concern the scope that a democratic system of government wishing to remain democratic should allow for microtargeting, manipulation, and amplification. The questions are important enough to warrant more than the usual knee-jerk responses, and they too should not be beyond the pale of civil conversation among twenty-first century civil libertarians.

IV. CONCLUSION

To appropriate a turn of phrase, the First Amendment is not a suicide pact. The mandate to preserve space for dissent, disagreement, and challenges to political and cultural consensus is vital, full stop. But the free speech imperative should not be interpreted to shelter the deliberate construction and fine-tuning of an information environment optimized to unravel the most basic preconditions for democratic self-government. It is platform functions and dysfunctions—rather than hierarchies and bottleneck effects, remediable failures of listener autonomy, or intermediary intentionality—that explain current threats to the anti-factionalism, anti-manipulation, and anti-authoritarianism interests. Platform functions and dysfunctions therefore should supply the frame for assessing constitutionally-required goodness of fit, and legislation appropriately tailored to the platform-based environment and its particular democratic failure modes should be correspondingly more likely to survive review.
The Russian attacks on voter registration databases preceding the 2016 Election “changed the narrative” about election security, highlighting a new avenue for adversarial foreign nation-states to interfere with the U.S.’s elections: the cyberattack.¹ This Technology Explainer will analyze how cyberattacks function in elections, which election systems are most vulnerable to these attacks, and how cyberattacks hurt election integrity.

II. CYBERATTACK METHODS IN ELECTIONS

A cyberattack is “an attempt to gain illegal access to a computer or computer system for the purpose of causing damage or harm.”² Some common cyberattack methods used on election systems are through denial-of-service (DoS) attacks, malware, Structured Query Language (SQL) injections, and phishing attacks.

DoS attacks interrupt or slow down access to a computer system for legitimate users by flooding the system with illegitimate traffic. In elections, computer systems vulnerable to DoS attacks include e-pollbooks, electronic voting machines, voter registration databases, and electronic auditing systems. The goal of a DoS attack is to render election technology unusable, thereby disrupting the election.

Malware is malicious software that introduces worms, spyware, viruses, Trojan horses, and ransomware to a system. Malware, like DoS attacks, can disrupt election technologies such as e-pollbooks, electronic voting machines, and electronic auditing systems. However, attackers can also use malware to attack these election systems to produce a specific desired result, such as manipulating vote counts on a voting machine or changing a person’s voting registration status on an e-pollbook.

A SQL injection “is a code injection technique that hackers can use to insert malicious SQL statements into [user-facing] input fields for execution by the underlying SQL database.” Attacker can use SQL injections to access and destroy data in voter registration databases.

A phishing attack is when an attacker, “masquerading as a trusted entity,” sends an email or text to a victim that prompts the victim into providing the attacker with sensitive information for that “trusted entity,” like a username and password. Attackers can use the sensitive information they obtain from phishing attacks to gain access to otherwise restricted voting systems.

---

4 Id.
6 Nat’l Acads. of Sci., Eng’g, & Med., supra note 3, at 86.
7 Id. at 86–87.
8 Id.
12 See Matthew Cole et al., Top-Secret NSA Report Details Russian Hacking Effort Days
III. Voting Systems Vulnerable to Cyberattacks

It is important to realize that any computerized voting system is vulnerable to cyberattacks. This Technology Explainer focuses on two voting systems: (1) voting machines and (2) voter registration databases. These technologies are integral to our elections and highly vulnerable to cyberattacks because they can be connected to the Internet.

A. Voting Machines

“Voting machines” are computers that cast and tabulate votes. There are multiple ways in which a cyberattacker can access a voting machine, but this paper will focus on methods that utilize the Internet.

First, although voting machines are not supposed to be connected to the Internet, they can be connected indirectly through election management computers, which are themselves connected to the Internet. Election management computers contain software and ballot definition files that are loaded onto voting machines using a cartridge or memory card. A cyberattacker could implant malware onto an election management computer via the Internet, and this malware could then be transferred to a voting machine by a cartridge or memory card. From this voting machine, the malware could further be transmitted to other voting machines through in-precinct local

---


16 Although generally not considered in the realm of cyberattacks, an attacker can also physically tamper with a voting machine, and studies have repeatedly shown how easy it is for an attacker with moderate levels of experience to corrupt a voting machine in person. See Matthew Blaze et al., DEF CON 27: Voting Machine Hacking Village 5 (2019), https://media.defcon.org/DEF%20CON%2027/voting-village-report-defcon27.pdf [https://perma.cc/88S3-MN4Z].

17 Manpearl, supra note 14, at 175; Nat’l Acads. of Sci, Eng’g, & Med., supra note 3, at 90–91; Schneider, supra note 10, at 255 (noting that “[a]lthough jurisdictions ‘should’ not connect those computers to a network or the Internet, no systematic efforts exist to ensure compliance with recommended security configurations.”).

18 Manpearl, supra note 14, at 175

19 Id.
networks and as memory cards are exchanged.\textsuperscript{20} Thus, through corrupting just one voting machine, the cyberattacker could corrupt the voting machines of an entire jurisdiction.\textsuperscript{21}

Second, some voting machines can be errantly connected to the Internet. For example, up until 2014, twenty percent of Virginia’s voting precincts were equipped with a wireless network to allow ballot programming and voter data exchange between voting machines.\textsuperscript{22} The Virginia State Board of Elections investigated these machines and discovered that wireless cards on the voting machine permitted “an external party to access the [machine] and modify the data [on the machine] without notice . . . an attacker could join the wireless ad-hoc network, record voting data or inject malicious [data].”\textsuperscript{23}

Regardless of how the cyberattack is carried out, once a voting machine is breached, the attackers can choose their desired impact on the machine from a parade of horribles. These impacts include flipping the vote the computer casts in favor of the attackers’ preferred candidate, destroying records required for auditing, or simply making the process of casting a vote more difficult.\textsuperscript{24}

If the voting machine does not have a paper audit trail (meaning it is not “software independent”\textsuperscript{25}), like Direct Recording Electronic (DRE) machines without a Voter Verified Paper Audit Trail (VVPAT), the only record of the vote is in the machine itself; in the event of tampering, there is no way to verify, audit, or recount the votes cast on that machine.\textsuperscript{26}

\textsuperscript{20}Id.; NAT’L ACADS. OF SCI, ENG’G, & MED., supra note 3, at 90.
\textsuperscript{21}Adam Aviv et al., Security Evaluation of ES&S Voting Machines and Election Management System, in EVT’08: PROCEEDINGS OF THE CONFERENCE ON ELECTRONIC VOTING TECHNOLOGY 4 (2008), https://www.usenix.org/legacy/event/evt08/tech/full_papers/aviv/aviv.pdf [https://perma.cc/87BW-Z3XH] (“a single circumvented piece of precinct hardware (such as a memory card returned from a precinct for vote tallying) can effectively ‘take over’ the county-wide back-end tally system, alter county-wide results reported in the current election, and then corrupt the installed firmware of additional precinct hardware in subsequent elections.”).
\textsuperscript{22}Manpearl, supra note 14, at 175–76.
\textsuperscript{23}Id at 176.
\textsuperscript{24}Id. at 179; NAT’L ACADS. OF SCI, ENG’G, & MED., supra note 3, at 85–86.
\textsuperscript{26}Schneider, supra note 10, at 254; Kimberly Breedon & Christopher A. Bryant, Counting the Votes: Electronic Voting Irregularities, Election Integrity, and Public Corruption, 49 U. MEM. L. REV. 979, 990, 993 (2019).
B. Voter Registration Databases

A voter registration database is a “single, uniform, official, centralized, interactive computerized statewide voter registration list . . . that contains the name and registration information of every legally registered voter in the state . . . .”27 Voter registration databases are considered “the most vulnerable part of U.S. election systems” because they are almost always directly connected to the Internet.28 Direct connection to the Internet makes voter registration databases susceptible to a variety of cyberattacks, including SQL injections, DoS attacks, and phishing attacks.29

Cyberattackers of voter registration databases can generally accomplish four objectives. First, they can make voters ineligible to vote by, for example, marking them as felons in a state where felons are not permitted to vote.30 In this way, the attacker could selectively disenfranchise voters to support their desired candidate.31 Second, the attacker can delete voter entries in the database prior to, or on, Election Day.32 In both of these scenarios, voters who cannot prove their registered status will be forced to cast provisional ballots, “leading to long lines, undermining faith in the fairness of an election, and creating a major administrative headache to accurately count votes after the polls closed.”33 Third, attackers could focus on vote-by-mail voting, changing address information of vote-by-mail voters or creating entries for voters that do not exist.34 Vote-by-mail voters who have had their addresses changed will not receive their ballots and may never vote at all.35 In the case of fictitious voters added to the database, it would be difficult for officials even

29 Schneider, supra note 10, at 252.
31 Manpearl, supra note 14, at 173–74.
35 Manpearl, supra note 14, at 174.
to recognize the problem without any in-person verification.36 Finally, if a DoS attack is employed and takes down the registration database, the e-pollbooks connected to this database would not be able to check in voters, creating widespread disruption.37

IV. WHY CYBERATTACKS HURT ELECTION INTEGRITY

Cyberattacks hurt election integrity through both actual and perceived tampering. The Supreme Court has consistently reiterated that the Constitution requires that each person’s vote is given “full and equal significance.”38 Both manipulating vote tallies and disrupting an individual’s ability to vote denies these rights entrusted by the Constitution.39

However, an election’s legitimacy does not only depend on accurate and equal vote counting; people must also believe that the election system in place upholds these qualities.40 Disruptions created by cyberattacks in even a small number of jurisdictions can lead to a loss of confidence in the integrity of the election as a whole.41 Doubts on vote-counting “crack the foundation on which the edifice of elections rests” and “may ultimately prove as destructive to the democratic processes as actual tampering.”42 Often casting doubt, rather than actually changing any election results, is the motive of a cyberattacker.43 An election without public confidence is not a legitimate election.44

V. CONCLUSION

The Russian attack from 2016 shows that the prospect of a cyberattack by foreign nation-states on election systems should not be considered a possibility, but an inevitability.45 Since 2016, the U.S.’s voting infrastructure is only becoming more dependent on voting technologies that are vulnerable to cyberattacks.46 One of Russia’s goals in the 2016 attack may have been to

36 NOREN & VANDEWALKER, supra note 33, at 16. For example, in 2012, hackers submitted online requests for thousands of vote-by-mail ballots in Florida, but thankfully the unusual activity was detected, and no disruption occurred. Id.
37 Breedon & Bryant, supra note 26, at 983–84.
38 Id.; Manpearl, supra note 14, at 174.
39 Breedon & Bryant, supra note 26, at 984.
40 Id. at 986.
41 Id. at 987.
42 Breedon & Bryant, supra note 26, at 982, 984.
43 Id. at 984.
44 Id. at 984.
45 Schneider, supra note 10, at 248–49 (“future attacks on American elections are inevitable.”).
learn about the U.S.’s election systems’ vulnerabilities for future attacks. The same cyberattackers behind the 2016 attacks also made successful attempts at penetrating the U.S.’s election systems before the 2018 midterms “and, by all accounts, will be back again in 2020.” This Technology Explainer has taken a first step in combating cyberattacks in our election systems by recognizing how they function in practice.

---


48 Thomas Hicks, Accessible and Secure: Improving Voter Confidence by Protecting the Right to Vote, in The Future of Election Administration 49, 49 (Mitchell Brown et al. eds., 2019).
DISTRIBUTED LEDGER TECHNOLOGY

Drew Diedrich*

CITE AS: 4 GEO. L. TECH. REV. 673 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ............................................................................................................. 673
II. WHAT IS DISTRIBUTED LEDGER TECHNOLOGY? .................................................. 675
III. DLT IN PRACTICE: BLOCKCHAIN AND DIRECTED ACYCLIC GRAPH .................. 676
IV. WHAT MAKES DISTRIBUTED LEDGER TECHNOLOGY APPEALING? ..... 678
   A. Benefits for Governments ..................................................................................... 679
   B. Benefits for Financial Institutions ..................................................................... 681
V. CHALLENGES FACING DLT ...................................................................................... 681
VI. CONCLUSION ............................................................................................................. 683

I. INTRODUCTION

Distributed Ledger Technology (DLT) is an emerging new spectrum of technologies focused on revolutionizing how data is stored.1 While often discussed in tandem with cryptocurrencies, not all DLTs are cryptocurrencies; instead the term DLT covers any approach to data storage across multiple ledgers.2 A ledger is a record of ownership and what is exchanged in transactions.3 The concept of a ledger is a foundational principle in modern

* Georgetown University Law Center, J.D. 2020; Cornell University, B.A. Government, 2017. Thank you to all the Georgetown Law Technology Review editors who helped with this piece.
2 WORLD BANK GRP., supra note 1.
economics with roots as far back as the 15th century. Traditionally, these records have been maintained in one centralized location, either with the government or with institutions such as banks and insurance companies. This centralized system requires all citizens to place significant amounts of trust in these entities to ensure accurate documentation of ownership within a ledger. Further, keeping the ledger in one place leaves it open to threats because it has a single point of failure for hackers and power outages. These centralized ledgers also create inefficiencies within transactions between parties.

With the growth of computing power and digital advancements, an alternative approach is now available: distributed ledger technology (DLT). With these new developments, a copy of a ledger can be changed in one node (a computing unit participating in distributed ledger network), and that copy can be distributed throughout a global network to all other nodes—in some cases within seconds of the change. In theory, DLT removes the need for trust required in our current economic system and replaces it with cryptography governing the distributed ledger. While this technology is in its infancy, there has been considerable investment in developing these technologies. DLT’s applications are growing and have significant implications within the legal environment. In order for practitioners to understand how to apply the law to DLT, they need to have an understanding of the technology.


5 WORLD BANK GRP., supra note 1, at 5; Systems Innovation, supra note 3.

6 WORLD BANK GRP., supra note 1, at 5.


8 U.K. GOVERNMENT CHIEF SCIENTIFIC ADVISER, supra note 1, at 5; WORLD BANK GRP., supra note 1, at VII.


10 Id.; U.K. Government Chief Scientific Adviser, supra note 1, at 5.

11 U.K. GOVERNMENT CHIEF SCIENTIFIC ADVISER, supra note 1, at 5.


13 REED SMITH LLP, supra note 3, at 1.
A distributed ledger is a traditional ledger that is recorded and stored in a number of data storage units. These data storage units are often referred to as nodes and can number in the hundreds or thousands across the world. While the underlying technology used to achieve a distributed ledger varies widely, all types of DLT use this concept of decentralized nodes. The nodes are connected through a shared software that allows them to communicate to each other through the network as they verify and record transactions. The ledger is stored within each node and when a change is made to one node’s ledger, a communication is sent out to other nodes to update the ledger.

The distributed ledger may be hosted on either a public or private network. Public networks are accessible to the general public and anyone may join the network to see the stored data. Becoming a node in a public DLT network can be as simple as downloading the software and connecting your computer to the Internet. An example of a public network is the Ethereum blockchain. In contrast, private networks require permission for entities such as users and nodes to join and may be preferred in sectors or industries that handle more sensitive information.

---

14 Id.
17 Heal, supra note 15.
18 Forex Academy, *How Is Distributed Ledger Technology Different From Blockchain?*, YOUTUBE (Oct. 18, 2019), https://www.youtube.com/watch?v=xFVjQv3cuYw&t=61s [https://perma.cc/85PM-BRW7].
20 Id. at 3.
21 WORLD BANK GRP., supra note 1, at 11.
22 Id.
23 FIN. INDUS. REGULATORY AUTH., supra note 19.
As previously mentioned, DLT is used in a wide spectrum of unique technologies.\textsuperscript{24} Two prime examples of this technology are blockchain and directed acyclic graphs (DAG). Blockchain is arguably the most famous example of DLT; it is the backbone infrastructure supporting some of the most recognizable cryptocurrencies.\textsuperscript{25} In a blockchain, transactions are bundled together into blocks which are added to the chain of previous blocks.\textsuperscript{26}

Ethereum is a good example to illustrate the process of verifying and storing information in blocks. Ethereum is a public programmable blockchain—which means it allows individuals to build applications using the blockchain as storage—and has a cryptocurrency called Ether that is used to pay for the computational power to run the applications.\textsuperscript{27} Within the Ethereum blockchain network, new transactions and updates to applications on the blockchain are periodically bundled into a “block” and broadcast to all nodes in the system.\textsuperscript{28} Nodes then race to validate the data in the block before it is added to the permanent ledger.\textsuperscript{29} If the new block is validated, the ledger is updated across all nodes and the nodes begin work on the next group of transactions.\textsuperscript{30} The confirmed block is then added to the linear chain of previous blocks and the updated chain of blocks is sent to all of the nodes in the system.\textsuperscript{31} Thus, the blockchain system works by distributing the data ledger across a network of nodes.\textsuperscript{32}

While blockchain may be the most famous example of distributed ledger technology, there are other competing approaches that also hold promise.\textsuperscript{33} An alternative approach using DLT is Directed Acyclic Graph (DAG) technology. Instead of storing transactions and data on one linear blockchain, with DAG-based technologies, each individual node

\textsuperscript{24} U.K. GOVERNMENT CHIEF SCIENTIFIC ADVISER, supra note 1, at 7.
\textsuperscript{26} Anwar, supra note 16.
\textsuperscript{28} COINTELEGRAPH, supra note 7.
\textsuperscript{29} Id.
\textsuperscript{30} Id.
\textsuperscript{32} Tellez-Merchan & Ricart, supra note 16.
\textsuperscript{33} Anwar, supra note 16.
independently stores and verifies transactions and data.\textsuperscript{34} Each node operates as a part of a one-way street, directing information to only ever pass in one direction (such as from older nodes to newer nodes), resulting in a structure known as a directed acyclic graph.\textsuperscript{35} This graph is acyclic in that one vertex cannot loop back around and reference itself.\textsuperscript{36}

Within a DAG-based network, users conduct transactions through their own nodes which appear on the network and are broadcast to the other nodes.\textsuperscript{37} Instead of using third parties to confirm the legitimacy of transactions on the ledger, each transaction itself either references or verifies previous


\textsuperscript{35} \textit{An Introduction to DAGs and How They Differ From Blockchain}, \textsc{fantom foundation}, (Jun. 20, 2018), https://medium.com/fantomfoundation/an-introduction-to-dags-and-how-they-differ-from-blockchains-a6f703462090 [https://perma.cc/YLC5-LPDA].

\textsuperscript{36} Malcolm Barrett, \textit{An Introduction to Directed Acyclic Graphs}, \textsc{inst. for sci. and mathematics of wirtschaftsuniversität wein} (Feb. 12, 2020), https://cran.r-project.org/web/packages/ggdag/vignettes/intro-to-dags.html [https://perma.cc/2B87-76T3]

transactions. Ultimately, because each transaction has to reference its parent transactions, the transactions can be traced back through the ledger’s history. The primary benefit for DAG over blockchain is scalability. Whereas blockchain limits the amount of transactions within a period of time, transactions on a DAG system eliminate the idea of blocks on the ledger and instead add transactions as the occur.

IV. WHAT MAKES DISTRIBUTED LEDGER TECHNOLOGY APPEALING?

DLT is beginning to expand across industries and sectors. Two of the most important benefits of DLT are increased efficiency and security. Because there is no middleman required to facilitate transactions, transactions can be conducted more quickly and efficiently. DLT can save businesses billions by accelerating the time it takes for a transaction between parties to settle and allowing transactions to be constantly processed every minute of the day as opposed to certain business hours dictated by third parties. Further, regarding supply-chains or transactions that require monitoring during the transaction, DLT reduces these monitoring costs by sharing verified information through the shared ledger.

DLT provides the means to maintain trust in a decentralized data system. Because the records are distributed and held in multiple locations,

38 CHURYUMOV, supra note 34; Meet the Tangle, supra note 37.
39 CHURYUMOV, supra note 34, at 5; Meet the Tangle, supra note 37.
42 BBVA, supra note 25.
43 REED SMITH LLP, supra note 3, at 11.
46 REED SMITH LLP, supra note 3, at 5.
47 Tellez-Merchan & Ricart, supra note 16.
there is, in theory, transparency between parties with access to the ledger. DLT also provides the means to maintain immutability within the ledger as decentralization ensures copies are stored in a number of locations.\textsuperscript{48} This decentralization protects the data within the nodes and decreases the risk of system failure.\textsuperscript{49} If a node or data center is compromised, it can be checked against the other nodes and expelled from the network.\textsuperscript{50} If someone were to infiltrate a node and change the data, the data would only be changed on that one node, not the rest of the network, and the network would reject that one compromised node.\textsuperscript{51} It is important to note that DLT does not solve all security problems. By its very nature, DLT creates more potential targets for nefarious hackers to attack.\textsuperscript{52} However, by distributing the ledger across a number of locations, the network becomes more resilient to certain types of attacks.\textsuperscript{53}

These dual benefits of transparency and immutability are available for potentially every application of DLT. Further, DLT may provide unique benefits for institutions that rely heavily on maintaining multiple accurate ledgers. Two such institutions are governments and financial institutions.

A. Benefits for Governments

Distributed ledger technologies may be used in the future to assist governments in collecting taxes, issuing passports, and maintaining government identification records.\textsuperscript{54} Governments are often targets for data breaches and current government storage presents a potential single point of failure because the ledgers are stored in one centralized data management...
system.\textsuperscript{55} DLT provides a security measure against malicious manipulation of the data because of it is stored across many different nodes.\textsuperscript{56} For certain types of cyberattacks such as a Distributed Denial of Service attack, a distributed ledger provides heightened security simply because the data is stored on a number of different computer systems.\textsuperscript{57} Also, citizen data maintained by the government would be harder to edit or erase on distributed ledgers without permission.\textsuperscript{58} Further, DLT can remove the need for individual agencies to have to ask citizens for the same information another agency already has by storing it across an accessible ledger as has been demonstrated by Estonia’s e-Estonia program.\textsuperscript{59}

DLT also holds promise for government identification programs and services.\textsuperscript{60} Digital identification offers a reliable and traceable record of transactions for vulnerable people who lack official identification.\textsuperscript{61} Especially for individuals “engaged in cross-border trade, seasonal migrants, or individuals displaced by conflict or humanitarian disasters,” DLT systems for storing personal data would provide stable digital identities that would be accessible anywhere in the world.\textsuperscript{62} These programs do face challenges to


\textsuperscript{57} Id.


\textsuperscript{59} Heller, supra note 58.


\textsuperscript{61} Id. While outside the scope of this paper, arguments have been made against these programs in that they overlook limited abilities for individuals, especially vulnerable populations, to access required technology so as to access their digital identification. For a brief overview of the risk, see WHITE ET AL., DIGITAL IDENTIFICATION: A KEY TO INCLUSIVE GROWTH 83 (Apr. 2019), https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20identification%20A%20key%20to%20inclusive%20growth/MGI-Digital-identification-Report.ashx [https://perma.cc/J3QW-78ZV].

\textsuperscript{62} Tellez-Merchan & Ricart, supra note 60.
adoption, including issues with “right-to-be-forgotten” laws and barriers to adoption due to unfamiliarity with the technology, and most of these projects remain in developmental stages.\textsuperscript{63} There is an inherent conflict between “right-to-be-forgotten” laws, which provides that citizens may request personal data be erased from organizational records, and an immutable ledger.\textsuperscript{64} Further, limited adoption due to the unfamiliarity of citizens with distributed ledger applications in a government context, most of these identification projects are in the developmental stage.\textsuperscript{65} Nevertheless, given that success in this area could mean globally accessible, reliable records of identification for people, DLT holds true promise in the realm of government identification.\textsuperscript{66}

B. Benefits for Financial Institutions

DLT also shows great promise for the financial industry. Banks are required to manage large amounts of data under strict governmental regulation.\textsuperscript{67} The application of distributed ledger technology to this data could cut down significantly on costs within the banking data systems.\textsuperscript{68} One study conducted by Accenture showed that investment banks could reduce their compliance costs between 30% and 50% by 2025 using DLT.\textsuperscript{69} Further, in regards to capital markets, DLT has the potential to reduce settlement times for securities transactions by facilitating the transfer of stocks and bonds simultaneously with the execution of a trade.\textsuperscript{70}

V. CHALLENGES FACING DLT

While DLT already been implemented in many fields, it faces several hurdles to widespread adoption. These hurdles include limited awareness by potential users about the technology, issues of scalability, unsettled laws surrounding DLT applications, and a general acceptance of current traditional technologies.

\textsuperscript{63} GALEN ET AL., supra note 58, at 21, 28.
\textsuperscript{64} Id.; Everything You Need to Know About the “Right To Be Forgotten,” GDPR.EU, https://gdpr.eu/right-to-be-forgotten/ [https://perma.cc/JW5E-KC4U].
\textsuperscript{65} GALEN ET AL., supra note 58, at 21, 28.
\textsuperscript{66} Tellez-Merchan & Ricart, supra note 60.
\textsuperscript{67} BBVA, supra note 25.
\textsuperscript{68} Id.
\textsuperscript{70} FIN. INDUS. REGULATORY AUTH., supra note 19, at 5–6; REED SMITH LLP, supra note 3, at 62.
In many applications of DLT, the technology is still in developmental stages. This lack of development reflects in part the overall youth of DLT but also reflects a lack of awareness on the part of potential users. Unaware potential users may not see the benefits presented by DLT to their current systems, which would limit their efforts to develop and adopt this technology in areas where it could be a benefit. Another primary challenge faced in the adoption of DLT is an issue of scalability, particularly with blockchain technologies. The Ethereum network processes only fifteen payments per second on average; that is significantly less than the average credit card company. While DAG-based solutions process transactions much faster than blockchain technologies, the most prominent examples of DAG-based DLT processing rates are still not as fast as conventional verification methods used by institutions such as credit card companies. Further, while DLT provides insulation from many types of security risks through its distributed nature, its distributed nature also exposes it more to potential risks outside the network. Bad actors have many more opportunities to try and attack the network through holes in software.

Another significant challenge specifically for financial institutions is unsettled and novel applications of law to DLT. As previously mentioned, for governmental identification through DLT, issues arise concerning an inherent conflict between “right-to-be-forgotten” laws and an immutable ledger. Another example of this problem is within financial institutions. Financial institutions are heavily regulated and are often changing.

---

71 FIN. INDUS. REGULATORY AUTH., supra note 19, at 10.
73 Id.
74 REED SMITH LLP, supra note 3, at 6.
77 FIN. INDUS. REGULATORY AUTH., supra note 19, at 11.
78 DESHPANDE ET AL., supra note 53, at 21.
80 Id.
Therefore, institutions and organizations attempting to implement DLT will have to navigate the application of traditional securities law to brand new technology.  

With potential application concerns comes the issue of satisfaction with current systems. If DLT does not provide significant benefits over current systems, businesses will ask why they should deal with the costs. Especially where existing systems are already efficient and adopted by users, DLT may face slow adoption. However, while this may be a barrier in established institutions, it may improve adoption efforts in new markets. New markets or areas without strong institutions may be the first to adopt DLT as its foundation because there is nothing for DLT to compete against.

VI. CONCLUSION

Distributed Ledger Technology has the ability to change our economy and the way we govern ourselves. Traditional institutions that rely on public trust can now be replaced with an open system that is checked and maintained by thousands of nodes all across the world. Current applications of DLT such as blockchain technology and directed acyclic graph technology show the potential of DLT; it improves security, preserves data integrity, and reduces costs. Governments and financial institutions in particular stand to benefit from these systems because of their heavy reliance on maintaining a large number of individual ledgers. In short, distributed ledger technology provides new solutions to many of the problems that we face in our modern world today. As the world increasingly accepts the digital age, the law and its practitioners will be required to understand distributed ledger technology from applications in government services and digital identification to financial services and beyond.

---

81 FIN. INDUS. REGULATORY AUTH., supra note 19, at 13.
82 Tellez-Merchan & Ricart, supra note 79.
83 Id.
84 Systems Innovation, supra note 3.
85 See Tellez-Merchan & Ricart, supra note 16.
BITCOIN OFF-CHAIN TRANSACTIONS: THEIR INVENTION AND USE

Michelle Mount*

CITE AS: 4 GEO. L. TECH. REV. 685 (2020)

TABLE OF CONTENTS

I. INTRODUCTION: WHAT IS AN OFF-CHAIN TRANSACTION, AND WHY IS IT IMPORTANT ................................................................. 685
II. A DIFFERENT BITCOIN MARKET THAN TEN YEARS AGO ................. 686
III. TRADITIONAL BITCOIN TRANSACTIONS VERSUS OFF-CHAIN TRANSACTIONS ........................................................................... 687
IV. MAIN METHODS OF OFF-CHAIN TRANSACTIONS .............................. 688
   A. Payment Channels ........................................................................ 689
   B. The Omnibus Wallet System ......................................................... 689
V. MAIN BENEFITS OF OFF-CHAIN TRANSACTIONS .............................. 690
   A. Safeguarding Privacy .................................................................. 690
   B. Bolstering Security Protections .................................................... 692
   C. Reducing Lengthy Settlement Times ............................................. 694
   D. Circumventing Volatile Transaction Fees .................................... 695
VI. MIXED REGULATORY RESPONSES TO OFF-CHAIN TRANSACTIONS .... 697
VII. CONCLUSION .................................................................................. 698

I. INTRODUCTION: WHAT IS AN OFF-CHAIN TRANSACTION, AND WHY IS IT IMPORTANT

Bitcoin off-chain transactions are transactions in bitcoin that are not recorded on the blockchain.¹ Today, most bitcoin transactions occur off-

---

* Georgetown University Law Center, J.D. Candidate 2020; Boston University, B.S. Business and Finance. A big thank you to the GLTR editors and to Professor Patrick McCarthy for his seminar class on Cryptocurrencies, Initial Coin Offerings and the Law.

chain.\(^2\) This dynamic runs counter to the quintessential principle espoused in the initial white paper conceptualizing bitcoin—that transactions would be immutably and reliably recorded on a public ledger.\(^3\) As regulators voice concerns over bitcoin’s divorce from its ledger technology, this explainer provides an understanding of this innovation’s implications on privacy, security, settlement speed, and usability.

II. A DIFFERENT BITCOIN MARKET THAN TEN YEARS AGO

Bitcoin now operates much differently than was envisioned in the initial white paper released by Santoshi Nakamoto in the aftermath of the 2008 global financial crisis. Bitcoin was designed to operate as a “purely peer-to-peer version of electronic cash [that] would allow online payments to be sent directly from one party to another without going through a financial institution.”\(^4\) Nakamoto proposed replacing financial intermediaries with blockchain technology,\(^5\) which uses a network of independently-run computers and cryptographic proofs to authenticate currency transfers.\(^6\) The system requires that bitcoin transactions be reported on the blockchain, which


\(^{3}\) SATOSHI NAKAMOTO, BITCOIN: A PEER-TO-PEER ELECTRONIC CASH SYSTEM 1, 8 (2008), https://bitcoin.org/bitcoin.pdf [https://perma.cc/HH9T-EHAS] (hereinafter BITCOIN WHITE PAPER) (proposing a “peer-to-peer network using proof-of-work to record a public history of transactions that quickly becomes computationally impractical for an attacker to change”).

\(^{4}\) Id. at 1.

\(^{5}\) EUROPEAN SEC. & MKTS. AUTH., DISCUSSION PAPER: THE DISTRIBUTED LEDGER TECHNOLOGY APPLIED TO SECURITIES MARKETS 8 (2016), https://www.esma.europa.eu/sites/default/files/library/2016-773_dp_dlt_0.pdf [https://perma.cc/SM3A-CTSV] (“Distributed ledgers’ and ‘Blockchain’ are often used interchangeably when discussing the technology. However, the Blockchain is a particular type of distributed ledger originally designed and used for Bitcoins.”).

\(^{6}\) BITCOIN WHITE PAPER, supra note 3, at 1 (“What is needed is an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party.”).
functions as the public immutable accounting ledger for the peer-to-peer network.  

Today’s bitcoin market departs from Nakamoto’s visionary concept in two ways. One, bitcoin is now more popularly used as a speculative investment product, not as an online payment vehicle. And two, most bitcoin transactions are facilitated by financial intermediaries. The result is a bitcoin market where most transactions are not reported to the blockchain.

III. TRADITIONAL BITCOIN TRANSACTIONS VERSUS OFF-CHAIN TRANSACTIONS

In a traditional bitcoin transaction, bitcoin is transferred from one digital wallet address to another. But there is no physical currency to transfer: bitcoin is a string of letters and number while a digital wallet is a public online address that can be accessed with a private key or passcode. Legally, the bitcoin is considered to be transferred after the majority of the network has

---

7 Id. at 8 (noting that the system prevents double spending through a “network using proof-of-work to record a public history of transactions that quickly becomes computationally impractical for an attacker to change”); id. at 3 (describing the six steps involved in reporting transactions to the blockchain); see also ABA DIGITAL ASSET PAPER, supra note 1, at 18–20 (giving a general description of blockchain technology).

8 See Mapping the Universe of Bitcoin’s 460 Million Addresses, CHAINALYSIS (Dec. 19, 2018), https://blog.chainalysis.com/reports/bitcoin-addresses [https://perma.cc/QD6K-X2MH] (“We estimate that on average only 20% of the bitcoin transaction value is economic, in that it is a final transfer between different people via economically relevant addresses.”).


10 See Hougan et al., supra note 2 (citing research finding that “95% of the [cryptocurrency trading] volume reported to popular data aggregators is either fake or wash trading” and is not reported on the blockchain); Can On-chain Data Help Us Spot Fake Exchange Trading Volumes? CHAINALYSIS BLOG (Nov. 15, 2019), https://blog.chainalysis.com/reports/fake-trade-volume-cryptocurrency-exchanges [https://perma.cc/7ZQK-QYTX] (finding the ratio of off-chain transactions to on-chain transactions is between 6:1 and (sometimes) 40,000:1); see also Kieran Smith, TIE Report Names and Shames Fake Volume Exchanges, BRAVE NEWCOIN (Mar. 26, 2019), https://bravenewcoin.com/insights/tie-report-names-and-shames-fake-volume-exchanges [https://perma.cc/4W87-3GKY].

11 BITCOIN WHITE PAPER, supra note 3, at 2 (defining bitcoin as a “chain of digital signatures. Each owner transfers the coin to the next by digitally signing a hash of the previous transaction and the public key of the next owner and adding these to the end of the coin.”); see Noelle Acheson, How to Store Your Bitcoin? COINDESK (Jan. 26, 2018), https://www.coindesk.com/learn/bitcoin-101/what-is-bitcoin [https://perma.cc/7ZJ8-E2PX] (explaining the mechanics of digital wallets).
updated the blockchain ledger to reflect that the specified bitcoin is associated with the receiver’s digital wallet.\textsuperscript{12}

In an off-chain bitcoin transaction, legal ownership of the bitcoin changes, yet because the bitcoin stays associated with the same digital wallet, the blockchain ledger does not update.\textsuperscript{13} To analogize to conventional currencies, a traditional transaction would be transferring money from one bank account to another, while an off-chain transaction would be putting money in a safe deposit box and giving someone the only key. In the second scenario, ownership of and access to the money changes, but the currency does not change locations, and the bank’s ledger does not update to reflect the ownership change. Bitcoin intermediaries have developed numerous off-chain methods to transfer legal ownership of bitcoin to circumvent the blockchain.

IV. MAIN METHODS OF OFF-CHAIN TRANSACTIONS

Off-chain transactions regularly occur through a payment channel or an omnibus wallet system. Custodians also effectuate off-chain bitcoin transactions for funds and exchanges.\textsuperscript{14} However, as the U.S. Securities & Exchange Commission (SEC) has noted, there are outstanding legal questions about the viability and industry standards of these services.\textsuperscript{15}

\textsuperscript{12} See David Mills, et al., \textit{Distributed Ledger Technology In Payments, Clearing, and Settlement} 13–14 (Bd. of Governors of the Fed. Reserve Sys., Discussion Series 2016-095), https://doi.org/10.17016/FEDS.2016.095 (describing how DTL or Blockchain technology transfers ownership of an asset); \textit{see also Bitcoin White Paper, supra} note 3, at 1 (“Digital signatures provide part of the solution . . . [and] [t]he network timestamps transactions by hashing them into an ongoing [block]chain[,] . . . forming a record that cannot be changed.”).

\textsuperscript{13} See ABA Digital Asset Paper, \textit{supra} note 1 and accompanying text.


\textsuperscript{15} Joint Statement, Div. Trading & Mkts., U.S. Sec. & Exch. Comm’n & Fin. Indus. Regulatory Auth., Joint Staff Statement on Broker-Dealer Custody of Digital Asset Securities (July 8, 2019), https://www.sec.gov/news/public-statement/joint-staff-statement-broker-dealer-custody-digital-asset-securities [https://perma.cc/7SU2-CMFJ] [hereinafter Joint Staff Statement] (“The Staffs acknowledge that market participants wishing to custody digital asset securities may find it challenging to comply with the broker-dealer financial responsibility rules. . . . As the market, infrastructure, and law applicable to digital asset securities continue to develop, the Staffs will continue their constructive engagement with market participants and to gather additional information.”).
A. Payment Channels

Payment channels are created through software programs that reference bitcoin’s software but avoid reporting every transaction immediately to the blockchain. The most popular bitcoin payment channel software is the Lightning Network. The Lightning Network, launched in beta form in March 2018, has grown exponentially, with its capacity increasing 62% between January 2019 and January 2020. Currently, the Lightning Network has the ability to transfer approximately $7 million in bitcoin. The Lightning Network allows parties to pre-allocate capital to a payment channel and keep accounts of transactions between themselves on their ledgers, with net amounts being transferred at a later time. Once the net amount is transferred, the final transaction is reported on the blockchain. Additionally, in December 2019, Bitfinex, one of the world’s largest cryptocurrency exchanges, integrated the network into its platform for easier transfers.

B. The Omnibus Wallet System

The omnibus wallet system is another off-chain mechanism used by cryptocurrency exchanges, which are currently responsible for 86% of the economically useful bitcoin wallet addresses. The omnibus wallet system facilitates exchanges’ private accounting ledgers. When a customer opts to trade on an exchange, their bitcoin is transferred from their digital wallet to an

---

16 Bitcoin’s software, called “Bitcoin Core,” runs on the network of computers that support the bitcoin blockchain; it maintains the rules for how cryptographic proofs are solved and how transactions are reported. Bitcoin Core, BITCOIN (Apr. 24, 2020), https://bitcoin.org/en/bitcoin-core/ [https://perma.cc/FC9S-DMVC].
17 Gudgeon et al., supra note 1, at 1–2.
19 Id.
22 Id.
24 Mapping the Universe of Bitcoin’s 460 Million Addresses, supra note 8.
25 ABA DIGITAL ASSET PAPER, supra note 1, at 33–34.
omnibus wallet that the exchange controls.\textsuperscript{26} When the customer purchases or sells bitcoins on the exchange, those coins stay in the omnibus wallet.\textsuperscript{27} The corresponding amount is debited or credited on the exchange’s internal ledger and reflected in the customer’s statement of account.\textsuperscript{28} Thus, these transfers are not reported on the blockchain.\textsuperscript{29} The SEC cited research showing that 95\% of the bitcoin transactions on cryptocurrency exchanges consist of these off-chain transactions.\textsuperscript{30}

V. MAIN BENEFITS OF OFF-CHAIN TRANSACTIONS

As bitcoin became a global store of value, innovative bitcoin businesses provided services to reduce the costs and risks of trading bitcoin.\textsuperscript{31} Bitcoin’s transformation into a rapidly traded asset class brought new challenges. For example, bitcoin’s public ledger of financial data created serious privacy concerns. Additionally, the lack of intermediaries meant that losses from theft or mistakes were almost irreversible. Further, bitcoin’s blockchain settlement time was far too long. And the fees to use bitcoin’s blockchain prohibited smaller transactions. Especially for intermediaries with additional financial and regulatory obligations, solving these problems was imperative.\textsuperscript{32} Off-chain transacting offered a popular way to address privacy concerns, provide more robust security, reduce costs, and increase transaction speed.

A. Safeguarding Privacy

The transparent nature of bitcoin’s public ledger creates privacy challenges.\textsuperscript{33} Many people believe that blockchain technology is somewhat anonymous because a person does not need to provide their legal identity to

\textsuperscript{26} Id.
\textsuperscript{27} Id.
\textsuperscript{28} Id.
\textsuperscript{29} Id.
\textsuperscript{31} See Matthew Hougan et al, supra note 2, at 4–6 (discussing the rise of bitcoin exchanges).
\textsuperscript{32} See infra notes 39–40 and accompanying text.
transfer bitcoin. But the details of that transaction are broadcast to the blockchain public ledger, resulting in a record of the amounts of bitcoin moving to different wallet addresses. Nakamoto forewarned that, by reviewing the transaction data, the wallet address can be connected with the associated transactions and that when amounts from two wallets are combined and spent in one transaction, the transactions from both wallets may be traced to one wallet owner.

As it turns out, gleaning real-world identities from merely looking at the blockchain is not terribly difficult. Regulatory agencies and cryptocurrency exchanges have increasingly relied on companies that map the blockchain and connect suspicious or illicit transactions to various bitcoin addresses. Chainalysis, one of the more popular providers of this service, claims to have already identified 147 million, or 86%, of the relevant bitcoin wallet addresses. One identity mapping method is a “dust attack.” In a “dust attack,” identity thieves send small amounts of bitcoin to digital wallets in an attempt to analyze the wallet owner’s spending patterns and uncover their identity. The potential for identity mapping engenders serious security concerns. Once a bitcoin holder’s identity is uncovered, the holder could be the target of phishing or physical attacks.

In contrast, off-chain transactions do not broadcast information on the public ledger. This alternative practice provides additional privacy that is particularly helpful to bitcoin intermediaries who are “engaged as a business

---

34 See id.; see also BITCOIN WHITE PAPER, supra note 3, at 6 (“The public can see that someone is sending an amount to someone else, but without information linking the transaction to anyone.”).

35 BITCOIN WHITE PAPER, supra note 3, at 6.

36 Id. (advising that “a new key pair [to a wallet address] should be used for each transaction to keep them from being linked to a common owner…Some linking is still unavoidable with multi-input transactions, which necessarily reveal that their inputs were owned by the same owner. The risk is that if the owner of a key is revealed, linking could reveal other transactions that belonged to the same owner.”). Because of bitcoin dust—small amounts of bitcoin left over from larger transactions—combining the amounts in two wallets is often necessary.


39 Mapping the Universe of Bitcoin’s 460 Million Addresses, supra note 8.


41 See supra note 1 and accompanying text.
in the exchange of virtual currency.\textsuperscript{42} These companies are required by the Bank Secrecy Act to collect their customers’ identification information for reporting purposes.\textsuperscript{43} As a result, these intermediaries carry additional privacy risk by holding customers’ personal information along with customers’ digital wallet information. In a cybersecurity breach, off-chain transactions can provide another layer of protection between their customers’ identities, private keys, and financial data.\textsuperscript{44}

B. Bolstering Security Protections

Bitcoin operates through a network of independently-run computers, so no governing third party can reverse fraudulent transactions.\textsuperscript{45} If a hacker gains access to a server that holds a user’s private key information, they can use the private key to gain control of the user’s digital wallet and irreversibly transfer the bitcoin.\textsuperscript{46} As Chainalysis advised in its 2019 report on crypto


\textsuperscript{45} See BITCOIN WHITE PAPER, supra note 3, at 1 (noting that by removing financial intermediaries “[c]ompletely non-reversible transactions” would be possible).

\textsuperscript{46} See ABA DIGITAL ASSET PAPER, supra note 1, at 20–21 (discussing the mechanisms of digital wallets); see also generally CHAINALYSIS, CRYPTO CRIME REPORT (2019), https://assets.website-files.com/5a9360f88433cb00018022c2/5c4f67ee7deb5948e2941fda_Chainalysis%20Januar
crime, “Hacking is on the rise partly because it works. It is hard to defend against given the scale of the adversaries. So, the stakes are high for exchanges and the cryptocurrency ecosystem more generally.” In 2018 alone, cryptocurrency hackers stole $1 billion worth of digital assets.

Additionally, bitcoin are susceptible to being lost and misplaced. Without a third-party intermediary, lost bitcoin passwords, or private keys, are not recoverable and there is no controller to reverse the transaction if a user inadvertently sends bitcoin to the wrong digital wallet address. A 2018 study estimated that more than 13.5% of all mined bitcoin had been lost. At the time of the writing of this article, that amounts to $142.2 billion.

In an off-chain transaction, bitcoin stays in the same digital wallet address, and a private key is not used. Therefore, private keys may be held off-line or in a manner that is less accessible to hackers, reducing the risk of theft. Additionally, when a cryptocurrency exchange initiates an off-chain transaction on its private accounting ledger, the exchange can reverse it. Especially for bitcoin businesses, reliance on off-chain transactions can mitigate the cost of mistakes in governance, operational structure, and network security.
C. Reducing Lengthy Settlement Times

Bitcoin transactions can take hours to settle. This delay is partly due to an encoded limitation in bitcoin’s software and partly due to market dynamics. Bitcoin’s software is limited to processing up to seven transactions per second. However, it takes around ten minutes for a group of transactions to be confirmed and reported in a block on the blockchain. If there is a disputed transaction, it may take six blocks (around one hour) before the network reaches a satisfactory consensus on the transaction’s validity. The processing speed also depends on the cost of the transaction fee paid. If a user opts to pay a cheaper transaction fee, they may only be guaranteed that the transaction is reported in one of the next six blocks. As a result, the clearing times for blockchain transactions could be up to two hours.

Off-chain transactions through exchanges and payment channels do not utilize the blockchain and, therefore, can be executed and settled immediately. These benefits are critical for cryptocurrency exchanges,

58 See Joseph Bonneau, How Long Does It Take for a Bitcoin Transaction to Be Confirmed, COINCENTER (Nov. 3, 2015), https://coincenter.org/entry/how-long-does-it-take-for-a-bitcoin-transaction-to-be-confirmed [https://perma.cc/M8TC-MSRC]; BITCOIN WHITE PAPER, supra note 3, at 7–8 (stating “[w]e now consider how long the recipient of a new transaction needs to wait before being sufficiently certain the sender can’t change the transaction. [Calculating] the probability the attacker could still catch up . . . we can see the probability drop off exponentially with z,” and finding that the probability is less than 0.1% after 5 blocks).
59 David Easley et al., From Mining to Markets: The Evolution of Bitcoin Transaction Fees, 134 J. FIN. ECON. 91, 95 (2019) (“Users submit transactions . . . they want verified and posted on the blockchain. They play a game in which they can chose to pay a fee to move up in the queue and thus reduce their waiting time, or they cannot pay a fee and experience a longer waiting time.”).
60 Bitcoin Transaction Fees, BILLFODL (Mar. 21, 2020, 1:20 PM), https://billfodl.com/pages/bitcoinfees [https://perma.cc/JA9H-HDRL] (showing fees to have a transaction mined in the next block, within three blocks, and within six blocks).
61 See id.
which can execute billions of bitcoin transactions a day.\textsuperscript{63} Even in one large transaction between two parties, an off-chain transaction may be needed to reduce counterparty and clearing risk.\textsuperscript{64} Unlike the central clearing systems in securities, commodities, and derivatives, bitcoin’s decentralized system means that there is no central counterparty to guarantee the settlement before the transaction clears. Third parties can guarantee off-chain transactions and provide instantaneous execution to reduce parties’ counterparty risk and currency risk during periods of price volatility.\textsuperscript{65}

D. Circumventing Volatile Transaction Fees

Using bitcoin has costs.\textsuperscript{66} The businesses running the computers behind the peer-to-peer nodes (often called miners) use considerable electricity and sophisticated computing equipment while expecting to earn a profit.\textsuperscript{67} Initially, the system rewards the miners for their work with new bitcoin.\textsuperscript{68} Yet, Nakamoto believed in a finite supply of bitcoin, and he programmed the bitcoin software to decrease the miners’ compensation as the number of transactions rises, so eventually the miners’ compensation would be zero.\textsuperscript{69} Nakamoto hypothesized that as bitcoin became more popular, users would voluntarily pay fees to the miners to expedite their transactions, the size of which would be dictated by market demand.\textsuperscript{70}

\textsuperscript{63} Top 100 Cryptocurrency Exchanges by Trade Volume, COINMARKETCAP, https://coinmarketcap.com/rankings/exchanges/ [https://perma.cc/F7X4-3RWB].
\textsuperscript{64} See ABA DIGITAL ASSET PAPER, supra note 1, at 33 (noting that exchanges—which use off-chain transactions—mitigate counterparty risk).
\textsuperscript{65} See DISTRIBUTED LEDGER TECHNOLOGY, supra note 50, at 5–6 (noting that real-time settlement or the functional equivalent—which takes place in off-chain transactions—can mitigate counterparty risk and decrease market inefficiencies).
\textsuperscript{67} Id.
\textsuperscript{68} Id.
\textsuperscript{69} Easley et al., supra note 59, at 94 (estimating that the miners’ programmed reward will reach zero around the year 2140).
\textsuperscript{70} See BITCOIN WHITE PAPER, supra note 3, at 4 (envisioning transaction fees as an additional incentive for miners).
That hypothesis has come to pass. The fees paid to miners to expedite bitcoin transactions have increased rapidly and can vary wildly. Before 2017, the average fee per transaction rarely reached $1. Then in late 2017, the average transaction fee touched $55; it has since retreated. The market dynamics behind transaction fees are complex and hard for companies to protect against through hedges or budgeting. Bitcoin transaction fees are fixed per transaction, unlike credit card fees, which are calculated as a percentage of the dollar value of the transaction. In periods of high transaction fees, smaller amounts of bitcoin may be worth less than the transaction fees. Thus, these amounts become effectively unspendable—trapped as “bitcoin dust.” One study estimated that at the height of transaction prices in 2017, almost $50 million’ worth of bitcoin became bitcoin dust.

Off-chain transactions may be the only way to spend bitcoin dust and avoid its accumulation. Because transacting off-chain does not utilize the blockchain, there are no bitcoin transaction fees, and thus bitcoin is not trapped as dust. Circumventing transaction fees is critical for businesses that require high-volume trading, such as bitcoin exchanges and high-frequency bitcoin traders. Also, payment channels can facilitate small transactions that would not have been possible on the blockchain due to transaction fees. Overall, off-chain transactions can reduce users’ exposure to transaction fees’

71 Easley et al., supra note 59, at 93 (with table panel A, showing that from 2013 to 2015, total transaction fees stayed around $2.3 million, but in 2016 bitcoin users were opting to pay $13.5 million in fees for expedited processing time).
73 Id.
74 See Bitcoin Hash Rate, Miner Margins Shift Back into Growth 3 DIAR, Mar. 4, 2019, at 2 https://diar.co/volume-3-issue-8/#2 [https://perma.cc/N769-RDL9] (discussing the complex interplay of miners’ revenues, cash flow, overall gross margin, and the bitcoin hash rate, which put smaller miners out of the market); see generally Easley et al., supra note 59 (analyzing the market economics of bitcoin mining and its impact on miners’ and bitcoin users’ decisions).
75 Bitcoin Transaction Fees, supra note 60; but see Dhruv Bansal, Bitcoin Data Science Part 3: Dust & Thermodynamics, UNCHAINEDCAPITAL (Dec. 18, 2018), https://www.unchained-capital.com/blog/dust-thermodynamics/ [https://perma.cc/6M6S-UYE9] (explaining how fees are more complicated to calculate when various containers of bitcoin are spent in a single transaction).
76 See Bansal, supra note 75.
77 Id.
78 Id.
79 Chavez-Dreyfuss, supra note 62.
80 Poon & Dryja, supra note 21, at 1–4.
variability and to unexpected liquidity crunches resulting from bitcoin’s dust dynamic.

VI. MIXED REGULATORY RESPONSES TO OFF-CHAIN TRANSACTIONS

Regulators are still navigating the implications of off-chain bitcoin innovations. While off-chain transactions have made the cryptocurrency significantly more usable, there are risks in bitcoin’s widespread divorce from its blockchain accounting functionality. Thus far, the regulatory consensus is mixed. The SEC has raised a concern that because unregulated off-chain transactions now account for a significant portion of the total bitcoin market, bitcoin’s price is susceptible to market manipulation.81 The Commodity Futures Trading Commission (CFTC) cautioned that it may be hard to “adequately assess the inherent risk of virtual currency contracts.”82 The SEC and the Financial Industry Regulatory Authority have cautioned that there are still ongoing questions around broker-dealer custody digital assets.83 While the SEC has thus far disapproved every attempt to launch a bitcoin-based ETP on national stock markets,84 the CFTC has approved a variety of bitcoin derivative products to launch under a heightened regulatory review process.85

81 See Order Disapproving a Proposed Rule Change, as Modified by Amendment No. 1, To Amend NYSE Arca Rule 8.201–E (Commodity-Based Trust Shares) and To List and Trade Shares of the United States Bitcoin and Treasury Investment Trust Under NYSE Arca Rule 8.201–E, 85 Fed. Reg. 12595, 12600–01 (Mar. 3, 2020) (discussing the types of possible fraud and manipulation from unregulated off-chain transactions and noting that this could affect the values that the ETF would be based on); id. at 12600 n.66 (“[S]tating that the sponsor of the proposed ETP presented an analysis of the bitcoin spot market that asserts that 95% of the spot market is dominated by fake and non-economic activity, such as washtrades.”); but see Comm’r Hester M. Peirce, Sec. & Exchange Comm’n, Dissenting Statement, In Response to Release No. 34-88284; File No. SR-NYSEArca-2019-39 (Feb. 26, 2020), https://www.sec.gov/news/public-statement/peirce-dissenting-statement-34-88284 [https://perma.cc/6KAB-STD4].


83 Joint Statement, supra note 15 (“Staffs have been engaged with industry participants regarding . . . custody solution[s] for digital asset securities. . . . The Staffs encourage and support innovation and look forward to continuing our dialogue . . . .”).

84 See sources cited supra note 81.

VII. Conclusion

Bitcoin innovation will likely continue. Other exchanges and custodial wallet services may follow Bitfinex’s lead by integrating the Lightning Network into their systems, redistributing more bitcoin into payment channels. If cryptocurrency exchanges move to facilitate high frequency trading, the number of possible off-chain transactions per second will increase. More numerous and rapid transactions could make tracing fraudulent payments more difficult. An influx of activity in cryptocurrency exchanges could also result in wallets holding more bitcoin dust, spurring the need for further innovation. Bitcoin intermediaries, security experts, and regulators must stay vigilant as the marketplace continues to evolve.

87 Chavez-Dreyfuss, supra note 62.
88 CRYPTO CRIME REPORT, supra note 46.
89 Poon & Dryja, supra note 21, at 54.
INNOVATIONS IN INTERNET VOTING SYSTEMS

Michelle Mount

CITE AS: 4 GEO. L. TECH. REV. 699 (2020)

TABLE OF CONTENTS

I. INTRODUCTION .............................................................................................................. 699
II. THE INVENTION OF INTERNET VOTING AND ITS INHERENT
    CHALLENGES .............................................................................................................. 700
III. THE HISTORICAL INTERNET VOTING SYSTEM USED IN THE UNITED
    STATES ......................................................................................................................... 701
IV. CURRENT CONFIGURATIONS OF NEW INTERNET VOTING SYSTEMS .... 702
    A. Estonia’s Fully Digitalized E-Government System................................. 703
    B. Scytl’s End-to-End Verifiable Voting System............................................. 705
    C. Russia’s Recent Pilot of an Ethereum Blockchain-backed Voting
        System.................................................................................................................... 707
    D. The United States’ Pilot of a Smartphone-Voting System Using
        Biometric Registration......................................................................................... 709
V. CONCLUSION .................................................................................................................. 710

I. INTRODUCTION

Election technology is democracy’s critical infrastructure and must
withstand manipulation from inside and outside forces. But in recent years,
contested recounts have raised procedural questions. Reports issued on
Russia’s attempts to influence the 2016 presidential election reveal other
system vulnerabilities.1 Security experts and the House of Representatives
have called for a return to paper ballots.2 But some states are exploring how

---

* Georgetown University Law Center, J.D. Candidate 2020; Boston University, B.S. Business
    and Finance. My sincerest gratitude goes out to the fabulous GLTR editors, without whom this
    piece would not have been possible.

1 S. SELECT COMM. ON INTELLIGENCE, 116TH CONG., RUSSIAN EFFORTS AGAINST ELECTION
    INFRASTRUCTURE WITH ADDITIONAL VIEW, S. REP. NO. 116-XX (2019),
    [https://perma.cc/L2KW-E42L].

Internet voting and advances in blockchain technology, facial recognition software, and cryptography could improve the election process.

This explainer (1) describes Internet voting and its inherent challenges, (2) analyzes the United States’ traditional Internet voting system, and (3) examines the benefits and risks of novel systems being implemented in the United States and around the world.

II. THE INVENTION OF INTERNET VOTING AND ITS INHERENT CHALLENGES

The late 1990s brought the cheap personal computing technology and widespread Internet access that fueled countries’ first experiments with Internet voting. Internet voting systems allow voters to cast their votes online, using a computer or mobile device, in a remote and unsupervised location. But this label applies to a broad range of solutions, and none have gained universal acceptance. The systems vary widely because the feasibility of each system depends on each jurisdiction’s Internet access, priorities, budget, laws, and election risk, as well as the digital literacy of its voters. Additionally, the election process has discrete phases: ballot distribution, voter identity verification, ballot casting, vote tallying, and vote auditing. Governments may choose to perform all or some of these phases over the Internet.

The technological challenges are formidable; some would say insurmountable. Election tampering is increasingly prevalent, and a...
networked voting system must withstand hacks of unprecedented sophistication and power.\textsuperscript{9} Some voters worry more about interference by their own government. Thus, an optimal voting system offers transparency so that a voter can verify that their vote was counted if they distrust the system’s operator.\textsuperscript{10} At the same time, the system must guarantee voter privacy to prevent bribery and voter intimidation.\textsuperscript{11} The cost of a manipulated election is enormous, and the coded nature of computer language makes subversion difficult to detect.\textsuperscript{12} Furthermore, advanced encryption techniques must be used to ensure that voters are not vulnerable to hacks and identity fraud.\textsuperscript{13}

III. THE HISTORICAL INTERNET VOTING SYSTEM USED IN THE UNITED STATES

Currently, Internet voting is only available for some overseas military service members and other select citizens\textsuperscript{14}: since 1986 states have been required to provide absentee ballots to overseas members of the military,\textsuperscript{15} but they are not required to accept those ballots electronically.\textsuperscript{16} The group is relatively small: in 2016 over 100,000 votes were cast online.\textsuperscript{17} There is still not a federally implemented secure voting solution for overseas military personnel and a twelve-year initiative to research such a system was repealed in 2015.\textsuperscript{18} The result is a piecemeal solution. Nineteen states and the District of Columbia accept some absentee ballots via email, and four states allow some voters to return ballots using a web-based portal.\textsuperscript{19}

While these technologies are comparatively inexpensive and easy to

\textsuperscript{9} Jefferson, supra note 7.
\textsuperscript{10} Id.
\textsuperscript{11} Id.
\textsuperscript{12} Id.
\textsuperscript{13} Id.
\textsuperscript{16} Internet Voting, VERIFIED VOTING (2018), https://www.verifiedvoting.org/resources/Internet-voting/#fn-45112-1 [https://perma.cc/T6QJ-V9DK] (discussing the reasons the MOVE Act is “notably silent on the subject of return of voted ballots”).
\textsuperscript{19} Electronic Transmission of Ballots, supra note 14.
administer, the risks are numerous. First, they compromise voter privacy—most states require online voters to waive their right to a secret ballot. Ballots submitted from another country move through networks of routers and forwarding agents before arriving at American election officials. Transmitted data can be viewed by national intelligence agencies. Second, ballots can be manipulated while in transit and detecting forged ballots is nearly impossible. Third, voters’ personal devices may be infected with malware, and the ballot files could carry malware into the election network. Security analysts caution that because these are risks inherent in the system architecture, they are unlikely to be remedied by encryption, firewalls, strong passwords, or voter signature checking.

IV. CURRENT CONFIGURATIONS OF NEW INTERNET VOTING SYSTEMS

The United States is not alone in its election security concerns. A report by the Canadian government found: “In 2018, half of all advanced democracies holding national elections had their democratic process targeted by cyber threat activity. This represents about a threefold increase since 2015.” These growing concerns about election legitimacy have sparked innovative solutions to address voter registration, system transparency, and ballot privacy.

This section provides a general introduction to four different types of Internet voting systems being used around the world. Part A discusses the world’s oldest nation-wide Internet voting system, which is infrastructure intensive and depends heavily on the legitimacy of the Estonian government for its legitimacy. In contrast, Part B reviews Scytl’s internet voting software, which offers end-to-end verifiability. Part C explores a blockchain-backed Internet voting system that was piloted in Russia. Finally, Part D looks at the

22 GREENHALGH ET AL., supra note 17, at 10.
24 Id.
25 Id.
26 Id.
27 Communications Security Establishment of Canada, supra note 8.
smartphone-accessible system with biometric registration which is being piloted in the U.S. Together, these examples highlight the key tradeoffs and considerations at play in the design and implementation of Internet voting systems.

A. Estonia’s Fully Digitalized E-Government System

Estonia has one of the most digitalized governments in the world, and in 2005, it became the first country to use Internet voting nationally. Its voting system draws on the country’s robust digital infrastructure. Most of the country’s public services are available online and accessible with citizen’s digital ID cards. One of core strengths of Estonia’s voting system is the pervasiveness of its national ID infrastructure and the sophistication of its cryptographic facilities.

Estonia’s unique system uses digital identification to provide authenticated access to its comprehensive e-government system, which allows citizens online access to almost all public services. The different information systems are connected through a secure Internet-based data exchange layer called the X-Road. The data is stored on servers run by the Estonian government that are located in both Estonia and Luxembourg. But the system’s critical registries are backed up on the private KSI blockchain. Because the KSI blockchain only stores coded references to the data, the raw

---

30 VASSIL, supra note 28, at 2.
31 See VASSIL, supra note 28, at 3–7; SPRINGALL ET AL., supra note 29, at 2, 10.
33 VASSIL, supra note 28, at 11.
35 Id.
data does not need to transfer out of Estonia’s system. This private
blockchain also provides a method of independent verification, which
safeguards against government manipulation.

Each Estonian with voting rights is issued an electronic national ID, or
Mobile-ID, which facilitates authentication, encryption, and digital signatures
when used within the voting software. Voters insert their digital ID into a
card reader that connects with their computer and downloads the voting
software. Once a vote is cast, a collector program verifies the vote and sends
it to a processor program to be anonymized for counting. An organizer
program then decrypts the votes using a private key and tallies them. The
votes are then audited and voters can verify that their vote was recorded
appropriately.

Estonia’s system is popular; in their 2019 European Parliament
election, 46.7% of voters cast their ballots online. But critics assert that the
system is not sufficiently transparent and that there are security issues.
Because the programs performing the key monitoring and authentication roles
are managed on government servers, citizens must have considerable trust in
the government running and protecting the elections. Additionally, the
startup and maintenance costs of Estonia’s digital ID infrastructure would be
costly. But proponents of digitalization argue that investing in a system

---

37 Id.
40 STATE ELECTORAL OFFICE OF ESTONIA, supra note 38, at 9.
41 Id.
42 Id.
44 SPRINGALL ET AL., supra note 29, at 2.
46 STATE ELECTORAL OFFICE OF ESTONIA, supra note 38, at 9.
47 SPRINGALL ET AL., supra note 29, at 2.
48 KINIRY ET AL., supra, note 5, at 113 (“Voters could be issued cryptographic ID cards such
which provides an authenticated digital identity may reduce others costs, such as the economic cost of fraud or of personnel to oversee laborious recounts.49

B. Scytl’s End-to-End Verifiable Voting System50

Scytl is a Barcelona-based software company that has facilitated online voting in more than 42 countries.51 Most notably, Switzerland, a pioneer in the Internet voting space since 2000, retained Scytl in partnership with its online election facilitator and postal service company, Swiss Post.52 Scytl offers an end-to-end verifiable voting solution,53 which solves the intrinsic problem of providing “public evidence from secret ballots.”54 End-to-end verifiability as the CAC cards issued to DoD personnel or like the national ID card of Estonia. … [But] the startup and maintenance costs will be very high. Voters would have to buy computers or devices that could read the cards, and they would almost certainly have to be useful for other online purposes besides just voting in order to justify the costs involved to both the government and the voter.”); see also Internet Voting in Estonia (iVote), JOINUP (Apr. 24, 2007), https://joinup.ec.europa.eu/collection/eparticipation-and-evoting/document/Internet-voting-estonia-ivote [https://perma.cc/RQC8-ZYQD] (“[Had there] not been a national population register or an authentication system using ID cards, the Internet elections in Estonia would have been very costly.”).

49 Internet Voting in Estonia, supra note 47 (noting that paper ballots are the most expensive voting method for those abroad and that replacing them with electronic votes will reduce costs); see also Jane Susskind, Decrypting Democracy: Incentivizing Blockchain Voting Technology for an Improved Election System, 54 SAN DIEGO L. REV. 785, 799 (2017) (describing how trustless systems in theory reduce transaction costs).

50 As this issue was going to publication, Scytl entered bankruptcy proceedings. The outcome of those proceedings is unclear, but the underlying technology remains relevant. This Technology Explainer will therefore refer to Scytl as an ongoing project. For news stories about Scytl’s bankruptcy, see Sonia Fenazzi, Swiss Post Set to Re-Launch Its E-Voting System, SWISS INFO (June 9, 2020, 10:45 AM), https://www.swissinfo.ch/eng/swiss-post-set-to-relaunch-its-e-voting-system/45820842 [https://perma.cc/P7PY-9UWR]; El Juze Abre La Venta De Scytl y Espera Ofertas Por La Empresa Hasta El 22 De Junio, LA VANGUARDIA (June 8, 2020, 10:42 AM), https://www.lavanguardia.com/economia/20200607/481657532094/scytl-subasta-unidad-productiva-venta-indaria-liquidacion.html [https://perma.cc/2NEA-HPT4].


54 See generally Matthew Bernhard et al., Public Evidence from Secret Ballots, in 10615
allows the voter to (1) check that their vote was recorded correctly, (2) check that their vote was included in the final tally, and (3) double-check that their vote was counted in the votes announced in the election outcome. This means that voters can have confidence in the system without having to trust the election officials. This verification system is complex; it requires the system to use advanced encryption techniques while still allowing voters to access their data at each stage.

Scytl’s system accomplishes end-to-end verifiability through a sequence of cryptographic measures. Once the vote is cast the ballot is immediately encrypted on voter’s device, instead of in transmission. Next, the ballot is sent to Scytl and privacy is ensured through use of cryptographic mixnets. Mixnets—mix networks—are generally used to make communication between sending and receiving networks untraceable. To do this, they shuffle messages and forward them to the next destination (possibly another node in the mix network) in a random order. Mixnets anonymize the voter’s identity by shuffling the data and breaking the original voting order in a process that re-encrypts and then decrypts the votes. Finally, voters can use the return codes they receive on their devices to check the public Bulletin Board and verify that their vote was properly recorded. Also during this stage, independent auditors or the media can observe and verify the vote counting and vote decryption process.

Scytl’s system, though, has notable drawbacks. One issue is that Scytl’s system does not address the problem of secure voter registration.
have issued one-use digital certificates or electronic ID cards to access the system.\textsuperscript{65} Additionally, some experts cite older studies of other end-to-end verifiable systems arguing that such systems are “complex and notoriously difficult to use,” and worry that this could disincentivize voters.\textsuperscript{66}

Finally, researchers found security flaws in Scytl’s software. The first of these flaws was a “cryptographic trapdoor” that would have allowed a malignant entity to insert or remove votes while votes are shuffled in the mixnet.\textsuperscript{67} Even though that first issue may have been corrected, this same group of researchers found an additional issue in the decryption process.\textsuperscript{68} Security developments are ongoing.\textsuperscript{69}

C. Russia’s Recent Pilot of an Ethereum Blockchain-backed Voting System

In 2019 Russia hosted one of the largest ever tests of a blockchain-
based Internet voting system in a binding national election during its parliamentary elections in Moscow.\(^70\) The system required voters to register online through their existing mos.ru account or apply in person with their passport.\(^71\) The mos.ru account is a government web portal where citizens can access certain services. Upon accessing the ballot, the voter was prompted to enter a confirmation code that they subsequently received via SMS at the mobile number connected to their account.\(^72\) The voter then had fifteen minutes to cast their vote. In theory, voters could verify their vote online.\(^73\) But a security assessment found that “the verifiability properties were not as strong as what could be hoped for from a blockchain-based ledger.”\(^74\)

The system used smart contracts on a specific permissioned Ethereum blockchain.\(^75\) After voting, the encrypted ballots are not stored in the web portal;\(^76\) they are immediately recorded on the blockchain as transactions, one transaction per ballot.\(^77\) Users are given enough information to identify their ballot was recorded on the blockchain, and the other ballots remain encrypted.\(^78\) This configuration was designed to allow a voter to access only their own vote to ensure it was properly recorded.\(^79\)

More than 10,000 people used this Internet voting system.\(^80\) Although the system did not use mixnets, security experts found that the encryption was significantly hard to break.\(^81\) Even so, the system could not guarantee ballot secrecy.\(^82\) This is because, during the election, the blockchain holding the encrypted voting data was still web-accessible for few hours.\(^83\) And at the end of the election day, the private key was sent to the web-accessible blockchain for verifiability purposes.\(^84\) Analysts were then able to obtain this private key


\(^{71}\) Id.

\(^{72}\) Id.

\(^{73}\) Id.


\(^{75}\) Id. at 2, 8. The Ethereum blockchain records data when it is authenticated by consensus of the network nodes.

\(^{76}\) Electronic Elections, supra note 65.

\(^{77}\) GAUDRY & GOLOVNEV, supra note 69, at 7.

\(^{78}\) Id. at 9.

\(^{79}\) Id. at 7.

\(^{80}\) Id. at 9.

\(^{81}\) Id.

\(^{82}\) Id. at 9–10.

\(^{83}\) Id. at 9.

\(^{84}\) Id.
and use it to decrypt 9,810 ballots. This prompted government concerns about expanding the system, and the losing candidate intends to challenge the election outcome in court.

D. The United States’ Pilot of a Smartphone-Voting System Using Biometric Registration

A new type of smartphone-accessible Internet voting system was piloted in West Virginia’s 2018 primary elections. This system can authenticate voters through biometrics and therefore does not require an alternative digital ID. After receiving their absentee ballot, voters are instructed to download Voatz’s mobile app, which requires them to scan in their state driver’s license or passport, take a moving selfie, and touch the fingerprint reader. Using facial recognition technology, Voatz verifies whether the selfie matches the government ID, and if it does, the ID holder is eligible to vote. The app registers the mobile device to the voter’s fingerprint. This connection ensures that a voter can vote only on one device and that a device can be used by only one voter. Once this linkage is established, it is encrypted and the identifying information is deleted. The system preserves ballot secrecy with end-to-end encryption and immutably records the votes on a multi-node Hyperledger-based permissioned blockchain.

So far Voatz has withstood attempted hacks and limited pilots have been successful. The West Virginia pilot included 147 military absentee

85 Id. at 9–10.
86 Elena Rozhkova et al., Tsifroi Ne Sklad’ibaetsia [The Numbers Don’t Add Up], KOMMERS. (Sept. 17, 2019), https://www.kommersant.ru/doc/4095101 [https://perma.cc/XLH6-2BAE].
89 See id. (explaining how voters can register and vote).
91 Id.
92 Id.
93 Id.
voters. In Denver, Colorado’s 2019 pilot for its municipal elections, 232 ballots were returned and counted. For the first time, Internet voting made available to local citizens with disabilities in Utah County, Utah’s 2019 pilot. And in October 2019, two counties in Oregon announced their intention to pilot the system for overseas military voters.

V. CONCLUSION

There is no one-size-fits-all Internet voting solution. Some robust systems may be too costly. Skeptical voters may demand more verifiability. Transparent blockchain systems may not be able to guarantee privacy. And managing the voter registration process and data security continues to be challenging. Regardless, the potential gains in election legitimacy and security motivate governments to persist in testing solutions. And in the age of online-banking and shopping, voters are growing increasingly irritated by long voting lines and their questionably disparate socio-economic impact on voter turnout. Under mounting pressure from governments and voters, the innovation in Internet voting is likely to continue.

---

96 MOORE & SAWHNEY, supra note 85, at 6.
97 NAT’L CYBERSECURITY CTR., supra note 89, at 8.
100 Bill Hewitt, Online Voting and Democracy in the Digital Age, CONSUMER REP. (May 17, 2016), https://www.consumerreports.org/online-voting/online-voting-democracy-in-the-digital-age/ [https://perma.cc/24MA-JA4C] (“A 2016 survey of voting-age Americans revealed that 33 percent would be more likely to vote if they could do it from an Internet-connected device like a smartphone.”).
EXPLAINING ALGORITHMIC DECISIONS

Gabriel Nicholas*

CITE AS: 4 GEO. L. TECH. REV. 711 (2020)

TABLE OF CONTENTS

I. INTRODUCTION ........................................................... 711
II. BACKGROUND ................................................................ 714
   A. What Is an Algorithm? ................................................. 714
   B. What Is an Explanation? ............................................. 715
   C. Why Do Explanations Matter? ................................. 716
III. EXPLANATIONS IN PRACTICE ........................................ 717
    A. Hypothetical Machine Learning Problem ...................... 717
    B. Understanding Data ................................................. 718
    C. Explaining Features ................................................. 719
    D. Explanations and Dimensionality ............................... 722
IV. EXPLAINING REAL WORLD SYSTEMS ............................... 726
    A. Difficulties in Real World Systems ......................... 726
    B. Model-Agnostic Explanations ............................... 727
V. CONCLUSION ................................................................. 729

I. INTRODUCTION

In 2006, when Netflix was just a DVD rental service, it offered a $1 million prize to the team that could improve Netflix’s movie recommendation algorithm by ten percent.1 The competition became an academic lightning rod—thousands of teams entered and their work produced dozens of academic

* Gabriel Nicholas is a Joint Research Fellow at the New York University School of Law Information Law Institute and the New York University Center for Cybersecurity as well as a Fellow at the Engelberg Center on Innovation Law & Policy. Thanks to Katherine Strandburg, Aaron Shapiro, Mark Verstraete, Salome Viljoen, and the Journal editors for their comments and guidance.

works. During the first year of the competition, several teams made significant headway; for example, AT&T’s team, KorBell, improved the algorithm by 8.43%. But soon, progress stalled. Teams snowballed in size as they pooled their efforts in attempts to gain a few fractions of a percentage point. Three years later, two teams finally passed the 10.00% improvement mark: BellKor’s Pragmatic Chaos, a hybrid team of KorBell and Big Chaos, and The Ensemble, a twenty-three-team super group.

How did the winning teams’ recommendation algorithms work? What enabled them to predict that a user would award a certain movie three stars as opposed to four? Competitors lacked incentives to optimize their solutions to give explainable predictions, and as such, their final submissions were so complex and inextricable that even the engineers who worked on them had trouble explaining the internal logic. BellKor Pragmatic Chaos alone incorporated 107 algorithms in its final submission. Besides, neither solution was ever used in its entirety—by the time the challenge was over, Netflix had shifted its focus to streaming, and the winning algorithms were too specific to DVD rentals to be repurposed.

While Netflix may have little moral imperative to explain how it derives its movie recommendations, greater transparency can be critical for high-impact decision-making systems. Users and auditors need to be able to interrogate algorithms that make decisions in areas such as hiring, foster care, and parole. Recently, the public has called for several popular consumer-facing platforms to better explain their underlying algorithms;

See, e.g., Zhou et al., Large Scale Collaborative Filtering for the Netflix Prize, in ALGORITHMIC ASPECTS IN INFORMATION MANAGEMENT 347 (2008).  
Id.  
Id.  
Id.  
Id.  
Id.  
Id.  
Id.  
Netflix did in fact start surfacing some explanations for its algorithm by recommending micro-genres, such as “Imaginative Time Travel Movies from the 1980s.” See Amatriain & Basilico, supra note 7.  
systems like Google’s search, Facebook’s newsfeed, and YouTube’s recommendation system are facing public scrutiny for their potentially deleterious effects on society.

Explanations for algorithms are important in the legal context as well. American legislation, such as the Fair Credit Reporting Act and the Equal Credit Opportunity Act, mandates a minimal level of explanation from credit companies. The incorporation of algorithms into new judicial processes raises questions about algorithmic transparency in due process, adjudication, and anti-discrimination as well. The European Union has attempted to address these demands for accountability in the General Data Protection Regulation (GDPR) with a “right to explanation,” but the question of what constitutes an explanation is a hotbed for legal debate.

The goal of this Technology Explainer is to give a brief, practical overview of how computer scientists come to understand the ways algorithms make decisions. To this end, it pulls from the literature of a nascent academic field called “explainable artificial intelligence” (XAI). Part II defines some common XAI terms such as “algorithm” and “explanation” and provides some background on important concepts in explainable machine learning. Part III

---

19 Katherine Strandburg, Adjudicating with Inscrutable Decision Rules, in MACHINES WE TRUST: GETTING ALONG WITH ARTIFICIAL INTELLIGENCE (Marcello Pelillo & Teresa Scantamburlo eds., forthcoming 2020).
uses a machine learning hypothetical to explore what explanations look like and the factors that can muddle them. Part IV shows how these explanatory methods can fail in complex real-world systems and offers alternatives. Part V briefly concludes.

II. BACKGROUND

A. What Is an Algorithm?

Thomas Cormen’s canonical *Introduction to Algorithms* defines an algorithm as “any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output.” The term “well-defined” leaves significant wiggle room. It can refer to explicit rules that produce straightforward, deductive output, such as sorting and searching algorithms. It can also refer to inductive procedures that come up with their own rules by generalizing from examples. Algorithms that use the latter option are called *machine learning algorithms*.

Explanations for algorithms with well-defined outcomes, like sorting, are less interesting and useful than explanations with less well-defined outcomes, like machine learning. If an algorithm sorts a list of American presidents by their birthdays, the reason it puts William Howard Taft between Woodrow Wilson and Theodore Roosevelt is self-explanatory. In contrast, machine learning algorithms can be applied to problems with answers that are subjective, not yet known, or otherwise too difficult to determine with just human-written rules. These outputs are called *predictions*, and as such, they can be questioned, justified, and, of course, explained.

Machine learning uses one algorithm, a *learner*, to output another algorithm that makes predictions, a *model*. The learner reads in data as a set of numerical features, infers rules about those features that predict the desired value, and outputs a model that embodies those rules. It is like a scientific model—an imperfect set of rules based on multiple factors and observations, like the theory of gravity. However, unlike gravity, the rules of an algorithmic model can be inscrutable, even to the person who wrote the learner.

---

26 Géron, supra note 24, at 18. What is described here is actually just one type of machine learning: “supervised learning.” There are other types of machine learning (e.g., unsupervised, reinforcement) that work differently and are used for different tasks. See id. at 8. Supervised learning is the prototypical type of machine learning algorithm and is by and large the focus of this Technology Explainer.
27 Id.
B. What Is an Explanation?

The relatively young field of XAI has exploded in the past few years. One study found that there are 289 core papers on the topic and 12,412 citing papers, coming from a wide intersection of fields such as cognitive psychology, computer science, business, and medicine.  

Given the sudden, interdisciplinary interest in XAI, there is much disagreement over nomenclature in the field, particularly around the terms “interpretability” and “explainability.” Some scholars use the terms interchangeably, while others regard the terms to refer to antithetical approaches. Still others define interpretability as an attribute or type of explainability. Although many authors believe that explainability is a latent, immeasurable property, a few human–computer interaction (HCI) researchers have attempted to measure it directly to compare different algorithms.

Of the many definitions of explainability and interpretability, a majority refer to the concept of human understanding. This commonality is best captured by Doshi-Velez and Kim’s definition of explainability: “the ability to present in understandable terms to a human.” Notably, this definition does not depend on any mathematical or epistemological properties, nor any strong normative claims.

Explainability is as much about the human capacity to understand a model as it is about a model’s capacity to be understood. Too many variables

35 Doshi-Velez & Kim, supra note 29, at 2.
or unintuitive relationships between those variables can make the model
difficult to conceptualize all at once. Implementation details can get in the
way as well: the size of a code base, the size of an engineering team, and the
links between different modules can all obscure how an algorithm comes to
its decisions. Explainability also depends on an explanation’s audience—a
comprehensible explanation for a subject matter expert might not be
understandable to the general public.

C. Why Do Explanations Matter?

If an algorithm is, according to Cormen’s definition, both “well
defined” and “computational,” it would seem to follow that its internal
reasoning should be readily apparent. In practice, this is not often the case.
Algorithms with opaque internal workings are commonly known as black
boxes, and, in many tasks, they can outperform simpler, more explainable
models in speed and correctness.

When an algorithm benefits from being explainable, it is often as a
proxy for other desiderata, like ethics or fairness. Explanations can reveal
when an algorithm uses socially unacceptable factors to make decisions. For
example, a college admissions algorithm that tries to optimize the percentage
of offers accepted may systematically reject applicants from Wyoming, if no
one from the state has ever accepted an admissions offer before. Moreover, a
college admissions algorithm may use other features as proxies for
undesirable, sensitive decision criteria, such as race, gender, or socioeconomic
status.

36 This idea is also called simulatability. See Slack, supra note 34, at 1.
37 Burrell, supra note 31, at 5.
38 Marko Robnik-Šikonja & Marko Bohanec, Perturbation-Based Explanations of
Predictions Models, in HUMAN AND MACHINE LEARNING 159 (Jianlong Zhou & Fang Chen
eds., 2018).
39 See, e.g., Karen Simonyan & Andrew Zisserman, Very Deep Convolutional Networks for
[perma.cc/Q3L3-DEEG].
40 See infra Parts III.D and IV.A for examples. Scholars have argued against this point, putting
forth a theory of “Rashomon sets” that states, since there are so many ways to write a machine
learning algorithm for a single problem, there must be an interpretable solution. This has yet
to be shown empirically. See generally Lesia Semenova & Cynthia Rudin, A Study in
Rashomon Curves and Volumes: A New Perspective on Generalization and Model Simplicity
[perma.cc/D942-H44E].
41 See Doshi-Velez & Kim supra note 29, at 2.
42 Finale Doshi-Velez & Mason Kortz, Accountability of AI Under the Law: The Role of
Explanation 9 (Berkman Klein Center for Internet & Society, 2017),
https://dash.harvard.edu/handle/1/34372584 [https://perma.cc/P555-U8MX].
Explanations may also be valuable in and of themselves by letting users learn from and improve their underlying decision-making process. Doctors can learn from medical diagnosis algorithms that explain their reasoning with representative cases. Conversely, doctors can correct medical diagnosis algorithms if they find they are optimizing for undesirable variables. In one case, an image recognition algorithm used to detect pneumonia in chest x-rays was found to instead predict which hospital the x-ray was taken at.

However, not all algorithmic systems benefit from being explainable. Finale Doshi-Velez and Beene Kim describe several scenarios where this is the case. First, there may be no significant consequences for incorrect outcomes such as for algorithms that find the dominant color in a painting. Second, the system may be sufficiently understood or validated in practice. Aircraft collision systems fall into this category—they work very well and are so straightforward that explanations are unnecessary so long as aircraft continue not to collide.

III. EXPLANATIONS IN PRACTICE

A. Hypothetical Machine Learning Problem

Imagine a model that tries to predict which college freshmen will successfully graduate. It could take, as input, data on a student and then output a true/false prediction about whether they will graduate. This kind of model is called a classifier because it categorizes the input data into a class—here, graduate=true or graduate=false. The input data might look as follows:


46 See Doshi-Velez & Kim, supra note 29, at 3.

47 Id.

48 The model may also output how confident it is in the prediction. See Géron, supra note 24, at 56.

49 See Pedro Domingos, A Few Useful Things to Know About Machine Learning, 55 COMMS. ASS’N FOR COMPUTING MACHINERY (ACM) 78, 79 (2012).
What would it look like create a classifier that predicts whether or not Warren will graduate? The first necessary ingredient is a historical data set of freshmen that includes whether or not they graduated (Figure 2). This data is called training data because it will be used to train the algorithm to perform the desired prediction task.

```
[{
  name: "Ralph", height_inchess: 75, gender: "male", age: 18y144d,
  sat_score: 1420, graduate: true},
{
  name: "Gina", height_inchess: 64, gender: "female", age: 17y292d,
  sat_score: 1280, graduate: true},
{
  name: "Alice", height_inchess: 68, gender: "female", age: 18y2d,
  sat_score: 1300, graduate: false},
...
]
```

*Figure 2: Training data*

Then, that data needs to be run through a learner. The learner will take as input the historical freshmen data set and output a classifier that can make predictions on unknown values, like Warren. The type of learner used will determine the type of model produced. A few types of learners will be explored in Section D.

B. Understanding Data

Before examining the underlying math of a specific algorithm, it is important to understand where the training data comes from. Any explanation for an algorithmic decision will have to relate back to the original training data, so understanding the origins and idiosyncrasies of that data is crucial.

Data can either be repurposed from existing data sets or it can be generated or gathered expressly for a specific algorithm’s use. These concepts are respectively called readymade and custommade data, and each approach has its advantages and disadvantages. Readymade data is more often incomplete, missing data that could potentially improve the performance of the algorithm. In the aforementioned college graduation algorithm, that

---

50 See id.

51 MATTHEW J. SALGANIK, BIT BY BIT: SOCIAL RESEARCH IN THE DIGITAL AGE 8, 83 (2017) Salganik intentionally avoids two other common terms for these ideas, "found" and "designed" data, because they falsely imply that the prior is objective and not designed in any way. Id.
missing data might be high school GPA.\textsuperscript{52} Readymade data is often unstructured and weighed down with extraneous information.\textsuperscript{53} For example, the college graduation data might include repeat students that need to be removed, or staff and faculty data that need to be filtered out using another data set.

Custommade data can often reflect a non-representative sample, resulting in an algorithm that is difficult to generalize.\textsuperscript{54} If the college graduation data was generated using surveys, perhaps there might be confounding traits that skew who chose to respond. The sample may not represent the actual population or may misrepresent certain subpopulations. Additionally, the factors relevant to graduation may be different when the algorithm is run versus when the data was collected.\textsuperscript{55} These scenarios can limit the situations a given algorithm can be applied.

The way an algorithm operationalizes its theoretical concepts also bears on its explanations.\textsuperscript{56} An algorithm that optimizes for some abstract concept—such as happiness or intelligence—needs a quantifiable definition of that concept. Even seemingly straightforward concepts like whether or not someone has graduated are not obviously defined. Does it count if they graduated in more than four years? How are transfer students handled? Sometimes, especially in readymade data, the answers to these questions are unavailable.

C. Explaining Features

Machine learning engineers can tinker with how the learner comes up with the classifier to improve its predictive ability. The main way engineers do this is through feature engineering, experimenting with which input data gets used and how it is represented. In some situations, good feature engineering can improve a machine learning algorithm even more so than a larger training data set.\textsuperscript{57} Part of feature engineering is processing raw data (e.g., Figure 2) into numerical data (e.g., Figure 3). Machine learning data

\textsuperscript{52} Id. at 24.
\textsuperscript{53} Id. at 37. This is called “dirty” data, as opposed to “clean” data.
\textsuperscript{54} Id. at 29.
\textsuperscript{56} GÉRON, supra note 24, at 25.
\textsuperscript{57} For example, in Facebook’s click through rate algorithm, feature engineering was found to affect the prediction more than sampling. See Xinran He et al., Practical Lessons from Predicting Clicks on Ads at Facebook, in PROCEEDINGS OF THE EIGHTH INT’L WORKSHOP ON DATA MINING FOR ONLINE ADVERTISING 1 (2014), https://quinonero.net/Publications/predicting-clicks-facebook.pdf [https://perma.cc/Q73E-6XUQ].
must at some level be represented numerically because the learner produces a classifier by shifting around the weights of different features or combinations therein. This can be challenging because not all information is equally amenable to representation as computer-processable data.

Feature engineering is more than converting strings to numbers. In the same way that words used to represent an idea are inexact and subjective, so too are the derived features an engineer chooses to represent data. Take the following example of how an engineer could choose to represent the raw data in Figure 2. The “sat_score” property is entirely dropped while the name is encoded in the feature of whether or not it starts with a “G.” Any internal explanation for how the classifier comes up with its decisions will derive from these features or their interactions.

```
[    
    {height_inches: 75, is_male: 1, age_in_years: 18.39, 
     name_starts_with_g: 0}, {graduate: 1}],
    
    {height_inches: 64, is_male: 0, age_in_years: 17.80, 
     name_starts_with_g: 1}, {graduate: 1}],
    
    {height_inches: 68, is_male: 0, age_in_years: 18.01, 
     name_starts_with_g: 0}, {graduate: 0}],
...
```

Figure 3: Featurized data

The field of XAI delineates between two different types of explanations: local explanations and global explanations. Local explanations interpret how an algorithm handles a specific input and how deviations to that input might affect the output. If test data was run through the classifier generated from the data in Figure 3, a local explanation might be: “The model predicted Warren will not graduate, but it would have if he were seventy days older.”

---

58 Even categorical data, such as the color of someone’s eyes, needs to be converted into a numerical representation. For example, {eyes_blue:1, eyes_brown:0, eyes_green:1}. Note how true/false values are represented as 1/0.

59 See e.g., Cynthia Rudin et al., A Process for Predicting Manhole Events in Manhattan, 80 MACHINE LEARNING 1, 8, 17 (2010), https://dspace.mit.edu/handle/1721.1/57432 [https://perma.cc/95V9-KTE6] (showing that creating a computer-processable data set requires close, sometimes difficult collaboration between scientists and domain experts).

60 Obviously, this engineer has little domain knowledge about what factors might predict graduation.

61 Slack, supra note 34, at 1.

62 Id.

63 The second part of this explanation, where gives the conditions needed to change the prediction, can be considered a “contrastive explanation.” See Brent Mittelstadt et al.,
entire model and all its decision paths.\textsuperscript{64} A global explanation for the college graduation classifier might be: “For the most part, the model predicts tall males whose names start with ‘G’ and females who enter school at over 6600 days old will graduate.”

Just because an explanation is intelligible, does not mean its line of reasoning is intuitive. A model’s internal reasoning is based on correlations in the data\textsuperscript{65} (whether or not a student’s name starts with “G” does not cause them to graduate), which can be particularly unintuitive.\textsuperscript{66} In the words of Professor Paul Ohm: “We are embarking on the age of the impossible-to-understand reason, when marketers will know which style of shoe to advertise to us online based on the type of fruit we most often eat for breakfast.”\textsuperscript{67} However, the features themselves are not always as intelligible as the type of fruit we eat for breakfast.\textsuperscript{68} Feature engineering is a discipline unto itself,\textsuperscript{69} and machine learning practitioners will sometimes use convoluted statistical methods such as clustering to create new, performance-improving features.\textsuperscript{70}

When features are intelligible, they can still be difficult to conceptualize. Additionally, explanations can sometimes be difficult to visualize.\textsuperscript{71} Other times there are too many features involved. Studies show that people can only keep seven plus or minus two cognitive entities in their head at once,\textsuperscript{72} and the features used by algorithms are no exception.

\textsuperscript{64} Id. at 281.


\textsuperscript{68} Mittelstadt et al., supra note 22, at 281.

\textsuperscript{69} See, e.g., ALICE ZHENG & AMANDA CASARI, FEATURE ENGINEERING FOR MACHINE LEARNING: PRINCIPLES AND TECHNIQUES FOR DATA SCIENTISTS (2018).


\textsuperscript{71} Mittelstadt et al., supra note 22, at 281.

\textsuperscript{72} See George A. Miller, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, 101 PSYCHOL. REV. 343 (1994).
D. Explanations and Dimensionality

Even with only a few straight-forward features, not all algorithms are equally conducive to explanations. Whether a more or less explainable algorithm is used to solve a problem often has to do with the shape of the training data. This can be seen in practice by applying three different algorithms to the college graduation problem: linear regression, logistic regression, and support vector machines.

Assume the training data includes sixteen former college students, represented using only two features: height_inches and age_in_years. Each feature can be mapped as a dimension on a graph, arranged in a chart that looks as follows (Figure 4):

![Figure 4: College data, with a clear division](image)

A linear regression learner works by trying to determine a line that divides the data points by the variable that the algorithm will try to predict. (Here, whether graduate has a value of true or false.) The line in Figure 5 is the classifier. If graphing a new value, such as Warren from Figure 1, that is to the left of the line, the algorithm would predict graduate as false. On the right side, it would predict true. Linear classifiers can make for great local explanations—here the graph makes it easy to calculate the difference in height_inches or age_in_years necessary to change the prediction.
Figure 5: Linear regression of college data. The star represents the unknown data point of Warren, which here would be classified as “graduated.”

The visualization also makes for a good global explanation so long as the classifier only uses a few dimensions. Linear regression can theoretically use any number of dimensions, and each dimension gets its own feature. If a third feature is added, such as gender, the graph would become three dimensional, and the classifier lines a two-dimensional plane. However, adding more dimensions than this would make the classifier difficult to visualize, as once there are too many dimensions (e.g., six, twelve, or a thousand), humans cannot conceptualize the model all at once.73

Logistic regression is similar to linear regression except it can draw a curved line instead of a straight one. Logistic regression could work well for a data set that looks as follows (Figure 8):

---

73 See id.; see also Slack, supra note 34, at 5.
Figure 6: Logistic regression of college data

The global explainability is roughly the same between linear and logistic regression because it is just as easy to visualize. But the local explainability becomes a little more unintuitive because the relationship between graduate and the other two variables is now non-monotonic.\textsuperscript{74}

Explanations can be even more difficult for data that looks like this (Figure 7):

\textsuperscript{74} See generally Lipton, supra note 22, at 13.
Figure 7: College data with concentric circle shape, no straight line division

There is no simple line, straight or curved, that can consistently predict graduate, so linear regression will fail to give accurate predictions. However, by using an algorithm called a support vector machine (SVM), the shape of the data can be translated from two dimensions onto three and a prediction plane can be drawn between them:

Figure 8: Support vector machine of college freshman data
This third dimension is neither height, age, nor a third independent variable. Rather, it is a convoluted mix of height and age (represented by the function $f$), calculated specifically to maximize the distance between the two classes. This third dimension makes global and local explainability far more difficult because it is difficult to articulate.\(^{75}\)

Sometimes, instead of increasing the number of dimensions on the graph, a machine learning algorithm will want to reduce them by collapsing multiple features into the same dimension. If, for example, a data set uses two thousand features, reducing the number of dimensions used to represent them can make the model run better and faster.\(^{76}\) Dimensionality reduction, which could be applied to any of the three algorithms described above, usually works by merging correlated features\(^{77}\) or by clustering similar data points.\(^{78}\) Again, this makes for complex variables that cannot easily be decomposed into intelligible parts, resulting in an algorithm that is difficult to explain.

IV. EXPLAINING REAL WORLD SYSTEMS

A. Difficulties in Real World Systems

In recent years, advocates and academics have sought explanations regarding the algorithms that drive popular tech platforms.\(^{79}\) These systems tend to be particularly complex, in part because of their massive scale\(^{80}\) and in part because they are built by top machine learning specialists with the expertise to employ cutting edge, frequently difficult to explain algorithms.\(^{81}\) Often these systems employ multiple algorithms that compete against one


\(76\) GÉRON, supra note 24, at 259.

\(77\) Id. (describing merging correlated features through principal component analysis).

\(78\) Id. (describing clustering similar data points through t-distributed stochastic neighbor embedding (t-SNE)).


\(80\) See Burrell, supra note 31, at 5.

another or feed into each other, with the output of one becoming the input of another. These combinatorial tactics are called ensemble methods, which can make deriving explanations from an algorithmic system even more difficult.

YouTube’s recommendation algorithm uses such ensemble methods. Its model is trained on hundreds of billions of examples, and each data point can be broken into one billion parameters. The recommendation system reduces these parameters to a couple hundred features, some of which are easily explainable (i.e., whether a user is logged in, what they have previously searched) while others are not, because they derive from ensemble methods or some dimensionality reduction technique.

Outsiders to the recommendation algorithm do not know the features it uses, nor the values it optimizes for. YouTube claims this prediction value includes “click-through rate, watch time, and many other metrics that measure user engagement,” but what these “other metrics” are, and how each is weighted, is unknown. YouTube also says this metric is constantly A/B tested, which may be why different recommendations appear for a video after the page is refreshed. Despite YouTube publishing a paper on its architecture, opacity surrounding its recommendation algorithm’s features make it a black box.

B. Model-Agnostic Explanations

Thus far, this piece has talked about explanations as an intrinsic property of algorithms. Models have been more or less conducive to certain types of explanations based on either their mathematical machinations or the clarity of their dimensions. However, intrinsic explanations do not work for opaque algorithms like SVMs, and other popular algorithms such as deep neural nets. Ensemble methods further degrade the usefulness of the intrinsic approach.

However, explanations for algorithms can also be uncovered by observing their behavior and uncovering statistical patterns. This type of

---

83 Id. at 191.
84 Id. at 195.
85 Id. at 196.
86 Id. at 192.
87 Id.
explanation is called model-agnostic\textsuperscript{89} (or post-hoc)\textsuperscript{90} because it depends not on knowledge of the algorithm, but rather on experimentation with it, through trying inputs, observing outputs, and surmising patterns.\textsuperscript{91} The advantage of model-agnostic explanations is that they can work on models in the wild, even ones that are intentionally opaque.

Imagine that the previously discussed college graduation algorithm was sold by a third-party as a black box. A college administrator interested in a global explanation could try manipulating a single feature, such as height\_inches, to see how it affects the prediction.\textsuperscript{92} Doing this for multiple features can demonstrate the relative feature strength,\textsuperscript{93} and doing this for two features at the same time can show whether those features interact.\textsuperscript{94}

Another approach is to build a more transparent classifier in order to predict the output of the black box classifier.\textsuperscript{95} However, it is debatable whether or not a transparent algorithm's predictions of a black box algorithm count as actual knowledge of how the black box itself works because the internal reasonings for the two algorithms may differ and their predictions will unlikely be a perfect match. A third model-agnostic approach is explanation by example.\textsuperscript{96} Highlighting prototypical examples of a class, outliers, or pairs of close data points with different predictions can help human operators to infer the rules of an algorithm.\textsuperscript{97}

\begin{itemize}
\item \textsuperscript{89} Ribeiro et al., \emph{supra} note 32, at 1137.
\item \textsuperscript{90} Poursabzi-Sangdeh et al., \emph{supra} note 33, at 1.
\item \textsuperscript{91} Ribeiro et al., \emph{supra} note 32, at 1137.
\item \textsuperscript{92} See generally Ron Pearson, \textit{Interpreting Predictive Models Using Partial Dependence Plots}, \textsc{Comprehensive R Archive Network} (Feb. 21, 2020), \url{https://cran.r-project.org/web/packages/datarobot/vignettes/PartialDependence.html} [\url{https://perma.cc/DNG3-UNZ6}].
\item \textsuperscript{93} See generally Alex Goldstein et al., \textit{Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation}, 24 \textsc{J. Computational \& Graphical Stat.} 44 (2013).
\item \textsuperscript{94} See Giles Hooker, \textit{Discovering Additive Structure in Black Box Functions}, in \textsc{KDD '04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining} 575, 576 (2004), \url{http://faculty.bseb.cornell.edu/~hooker/VIN-kdd.pdf} [\url{https://perma.cc/296W-EXQ9}].
\item \textsuperscript{96} Been Kim et al., \textit{Examples are Not Enough, Learn to Criticize! Criticism for Interpretability}, in \textsc{NIPS 2016: Proc. 30th International Conference on Neural Information Processing Systems} 29 (2016), \url{https://papers.nips.cc/paper/6300-examples-are-not-enough-learn-to-criticize-criticism-for-interpretability.pdf} [\url{https://perma.cc/VQ5H-ADTV}].
\item \textsuperscript{97} See generally Agnar Aamodt \& Enric Plaza, \textit{Case-Based Reasoning: Foundational Issues},
\end{itemize}
There are some important limitations to the insights that can be gained from model-agnostic methods and their applications. First, they sometimes only work with knowledge of the features used to represent data. Imagine a user was given a blank text box to input into the college graduation algorithm. It is unlikely that user would be able to guess how to format the data. This can be somewhat mitigated by preprocessing, the means by which an application turns raw data, like a video, into algorithm usable data (also known as features). For example, users can upload videos to YouTube and see the recommendations made without knowing what features YouTube’s recommendation algorithm uses to represent the video.

Even with preprocessing, it can be difficult to figure out which variables to manipulate and anticipate what effects those manipulations might have. And it might be almost impossible to dissociate the real variables the model is using from the correlated ones. In this way, model-agnostic explanations cannot elucidate the real, internal reasons why a model makes its decisions. This may limit their capacity to help troubleshoot a model and, potentially, have legal implications for any explanation that requires a defined degree of certitude.

Model-agnostic explanations also require continuous access to the algorithm to submit data for testing. Software companies often limit user access to algorithmic systems under the auspices of security and protecting them from being gamed.98 YouTube’s recommendation algorithm uses upload limits and anti-web scraping technology to this end,99 plus it likely bolsters its recommendations with data not available to the public.

V. CONCLUSION

The public is increasingly asking for accountability from technological systems, and in turn, explanations for how algorithmic systems come up with decisions. This Technology Explainer has explored what stands in the way of those explanations, at least at a software level. But explainability in decision-making systems is not just an engineering problem. Algorithm creators will only optimize for explainability when they have a reason to. Laws such as the Fair Credit Reporting Act, Equal Credit Reporting Act, and the General Data

---


99 See e.g., robots.txt file for YouTube, YouTube, https://www.youtube.com/robots.txt [https://perma.cc/9NNB-4S6S].
Protection Regulation incentivize explanations to a certain extent in a select few areas, but other areas of the law, such as trade secret and copyright, incentivize the opposite.

Explanations are a starting point for accountability, not a destination. An explanation can describe how an algorithm makes decisions, but it does not come with built-in means of redress for when those explanations contradict larger societal values. This problem is not one for computer scientists alone; lawyers and policymakers must join the discussion around explaining algorithmic decisions to decide how and when these explanations can address societal harms.

100 Selbst & Barocas., supra note 66, at 1107.