Utah Law Review

Volume 2020 | Number 4 Article 3

7-2020

Information Hacking

Derek E. Bambauer
University of Arizona James E. Rogers College of Law

INFORMATION HACKING

Derek E. Bambauer[*]

Abstract

The 2016 U.S. presidential election is seen as a masterpiece of effective disinformation tactics. Commentators credit the Russian Federation with a set of targeted, effective information interventions that led to the surprise election of Republican candidate Donald Trump. On this account, Russia hacked not only America's voting systems, but also American voters, plying them with inaccurate data—especially on Internet platforms—that changed political views.

This Essay examines the 2016 election narrative through the lens of cybersecurity; it treats foreign efforts to influence the outcome as information hacking. It critically assesses unstated assumptions of the narrative, including whether these attacks can be replicated; the size of their effect; the role of key influencers in targeted groups; and the normative claim that citizens voted against their preferences. Next, the Essay offers examples of other successful information hacks and argues that these attacks have multiple, occasionally conflicting goals. It uses lessons from cybersecurity to analyze possible responses, including prevention, remediation, and education. Finally, it draws upon the security literature to propose quarantines for suspect information, protection of critical human infrastructure, and whitelists as tactics that defenders might usefully employ to counteract political disinformation efforts.

> *The great danger of lying is not that lies are untruths, and thus unreal, but that they become real in other people's minds.*
>
> Christine Leunens, *Caging Skies*

INTRODUCTION

The standard account of the 2016 U.S. presidential election is that the Russian Federation, at the direction of its President Vladimir Putin, hacked not only voting

[*] © 2020 Derek E. Bambauer. Professor of Law, University of Arizona James E. Rogers College of Law. I owe thanks for helpful suggestions and discussion to Erika George, Dan Hunter, RonNell Andersen Jones, Thinh Nguyen, Amelia Smith Rinehart, Alan Trammell, and the participants at the 2019 Utah Law Review Symposium. I welcome comments at <derekbambauer@email.arizona.edu>.

systems but voters.[1] A sophisticated Internet campaign directed carefully crafted political disinformation at parts of the American electorate,[2] resulting in the election of Donald Trump, a candidate with overt sympathies for Russian interests.[3] The narrowness of the Trump victory—he won three million fewer votes nationwide than Democratic candidate Hillary Clinton, and carried three crucial states by fewer than eighty thousand votes total[4]—and the erratic behavior of the resulting administration[5] have brought disinformation, particularly on social media platforms, under sharp scrutiny.[6] Commentators have analyzed the finely-honed targeting of disinformation to U.S. voters, along with evidence that Russia engaged in selective leaking of accurate information, to conclude that this subset of "fake news" changed the outcome of the election.[7] Some have gone so far as to pinpoint this interference

[1] *See, e.g.*, Jane Mayer, *How Russia Helped Swing the Election for Trump*, NEW YORKER (Sept. 24, 2018), https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump [https://perma.cc/L9MR-UA86].

[2] I define "disinformation" in this context to mean data that is known to be false and that is distributed with the intent to alter the political positions of voters or groups of voters. *See generally* Samantha Bradshaw & Philip N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation* (U. Oxford Computational Propaganda Research Project, Working Paper No. 2019.2, 2019), https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf [https://perma.cc/LU6N-XPHX] (analyzing the tools used by governments and political parties to manipulate social media).

[3] *See, e.g.*, Marshall Cohen, *25 Times Trump Was Soft on Russia*, CNN (Nov. 19, 2019, 1:48 PM), https://www.cnn.com/2019/11/17/politics/trump-soft-on-russia/index.html [https://perma.cc/87P5-RXZK].

[4] *See, e.g.*, Philip Bump, *Donald Trump Will Be President Thanks to 80,000 People in Three States*, WASH. POST (Dec. 1, 2016, 1:38 PM), https://www.washingtonpost.com/news/the-fix/wp/2016/12/01/donald-trump-will-be-president-thanks-to-80000-people-in-three-states/ [https://perma.cc/9MAP-XJ69]; Gregory Krieg, *It's Official: Clinton Swamps Trump in Popular Vote*, CNN (Dec. 22, 2016, 5:34 AM), https://www.cnn.com/2016/12/21/politics/donald-trump-hillary-clinton-popular-vote-final-count/index.html [https://perma.cc/Z9T5-W57J].

[5] *See, e.g.*, Daniel Lippman, *Trump Veterans See a Presidency Veering off the Rails*, POLITICO (Oct. 19, 2019), https://www.politico.com/news/2019/10/19/trump-white-house-staff-051393 [https://perma.cc/67F6-66ZM]. *See also generally, e.g.*, PETER BERGEN, TRUMP AND HIS GENERALS: THE COST OF CHAOS (2019) (focusing on the Trump administration's unorthodox actions in the foreign policy arena).

[6] *See* Bradshaw & Howard, *supra* note 2.

[7] *See generally, e.g.*, KATHLEEN HALL JAMIESON, CYBERWAR: HOW RUSSIAN HACKERS AND TROLLS HELPED ELECT A PRESIDENT: WHAT WE DON'T, CAN'T, AND DO KNOW (2018) (analyzing the effect that Russian hacking and social media messaging had on the 2016 presidential election); Young Mie Kim et al., *The Stealth Media? Groups and Targets Behind Divisive Issue Campaigns on Facebook*, 35 POL. COMM. 515 (2018) (discussing the use of digital media by anonymous political campaigns to affect the 2016 U.S. elections). The U.S. is not the only country facing these challenges. *See, e.g.*, Emilio Ferrara, *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election*, 22 FIRST

as the beginning of the end of U.S. superpower status and the relative Pax Americana it generated since the end of the Cold War.[8]

There is little if any doubt about Russian intentions.[9] The country's security service calculated, correctly, that a Clinton-led administration would pose far more opposition to Russian strategic interests than a Trump-led one would.[10] Investigations such as those by Special Counsel Robert Mueller uncovered compelling evidence of widespread Russian attempts to sway voters.[11] And, the outcome was likely better than even the most optimistic predictions of electoral interference; the Trump administration has consistently supported Russian interests even at significant costs to putative allies, such as NATO (North Atlantic Treaty Organization) member states.[12] The conventional wisdom is that Russia executed a strategic masterpiece.

Why, then, were the Russians surprised by the Trump victory?[13]

The hacking analogy can help us to answer this question. There are many ways to find and exploit vulnerabilities—bugs—in information technology systems. Some attacks are precise and elegant, such as the Stuxnet cyberweapon used to damage the centrifuges in Iran's uranium enrichment facility at Natanz. Stuxnet targeted specific

MONDAY 1–2 (2017), https://firstmonday.org/ojs/index.php/fm/article/view/8005/6516 [https://perma.cc/DZ7G-6DXS].

[8] *See, e.g.*, Robert Kagan, *Trump Marks the End of America as World's 'Indispensable Nation*,' FIN. TIMES (Nov. 19, 2016), https://www.ft.com/content/782381b6-ad91-11e6-ba7d-76378e4fef24 [https://perma.cc/J8QD-TK7F].

[9] The 2016 attack is merely the latest in a long history of interference. *See* Morten Bay, *Fiona Hill's Story of Russian Disinformation Sounds Very Familiar*, SLATE (Nov. 22, 2019, 1:38 PM), https://slate.com/technology/2019/11/fiona-hill-russia-disinformation-testimony-history.html [https://perma.cc/W8MC-6Y2W]; Sean Illing, *"Flood the Zone with Shit": How Misinformation Overwhelmed Our Democracy*, VOX (Feb. 6, 2020, 9:27 AM), https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation [https://perma.cc/JFQ5-ZPYY].

[10] *See* Chad Day & Eric Tucker, *Mueller Revealed His Trump-Russia Story in Plain View*, ASSOCIATED PRESS (Mar. 22, 2019), https://apnews.com/3c4bc6e9aa6c4fb18bc9603fb082af65 [https://perma.cc/B8LB-LS2J].

[11] *See id. See generally also* ROBERT S. MUELLER, III, U.S. DEP'T OF JUSTICE, REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION (Mar. 2019), https://www.justice.gov/storage/report.pdf [https://perma.cc/82GF-9HFY] (detailing the evidence of Russian interference in the 2016 presidential election).

[12] *See, e.g.*, Reese Erlich, *Russia Is the Only Winner in Syria*, FOR. POL'Y (Oct. 30, 2019, 9:15 AM), https://foreignpolicy.com/2019/10/30/russia-is-the-only-winner-in-syria/ [https://perma.cc/N93G-JZS2]; Sebastian Sprenger, *Iran Fallout Deepens Rift Between America and Europe*, DEFENSE NEWS (Jan. 10, 2020), https://www.defensenews.com/global/europe/2020/01/10/iran-fallout-deepens-rift-between-washington-and-europe/ [https://perma.cc/N67C-4RMA].

[13] *See* Mark Hensch, *US Caught Russian Officials Cheering Trump Win: Report*, THE HILL (Jan. 5, 2017, 7:52 PM), https://thehill.com/policy/international/russia/312961-us-caught-russian-officials-cheering-trump-win-report [https://perma.cc/3V72-3TX4].

weaknesses in the Siemens programmable logic controllers and the Microsoft Windows print spooler service.[14] Some are random, brute force attacks, such as fuzzing, which bombards a system with data until one combination causes it to break.[15] People think Russia used the psychological equivalent of a zero-day attack[16] to influence American voters.[17] It is far more likely they engaged in social media fuzzing: trying a variety of information tactics,[18] and then being pleasantly surprised when some of them worked.[19] In all likelihood, no one in either the United States or Russia knows what vulnerability the attackers exploited or whether the hack can be repeated.[20] That may make the intervention more frightening. But hacking has much to teach about defending against exploits as well.

[14] *See* Gregg Keizer, *Microsoft Confirms It Missed Stuxnet Print Spooler 'Zero-Day,'* COMPUTERWORLD (Sept. 22, 2010, 2:57 PM), https://www.computerworld.com/article/2515799/microsoft-confirms-it-missed-stuxnet-print-spooler--zero-day-.html [https://perma.cc/7EMT-US2H]; David E. Sanger, *Obama Order Sped Up Wave of Cyberattacks Against Iran*, N.Y. TIMES (June 1, 2012), https://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html [https://perma.cc/WHM7-Y9Y5]; Kim Zetter, *An Unprecedented Look at Stuxnet, the World's First Digital Weapon*, WIRED (Nov. 3, 2014, 6:30 AM), https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/ [https://perma.cc/U28C-3XQF].

[15] *See* Andy Greenberg, *Hacker Lexicon: What Is Fuzzing?*, WIRED (June 2, 2016, 7:00 AM), https://www.wired.com/2016/06/hacker-lexicon-fuzzing/ [https://perma.cc/7VZQ-ZPBQ].

[16] *See generally* FireEye, *What Is a Zero-Day Exploit?*, https://www.fireeye.com/current-threats/what-is-a-zero-day-exploit.html [https://perma.cc/6EW5-MRMM] (explaining that a zero-day exploit is an unknown software or hardware flaw that can be exploited by attackers before a developer has an opportunity to fix the vulnerability).

[17] *See, e.g.*, JAMIESON, *supra* note 7; Massimo Calabresi, *Inside Russia's Social Media War on America*, TIME (May 18, 2017), https://time.com/4783932/inside-russia-social-media-war-america/ [https://perma.cc/4G2V-FSEQ]; Molly McKew, *Did Russia Affect the 2016 Election? It's Now Undeniable*, WIRED (Feb. 16, 2018, 10:25 PM), https://www.wired.com/story/did-russia-affect-the-2016-election-its-now-undeniable/ [https://perma.cc/8NUN-UJGP].

[18] *See* Abigail Abrams, *Here's What We Know So Far About Russia's 2016 Meddling*, TIME (Apr. 18, 2019), https://time.com/5565991/russia-influence-2016-election/ [https://perma.cc/T4DT-7EAD].

[19] *See generally* Peter Griffin, *Hacking the Human: Why Most Cybercrime Doesn't Involve Computer Hacking*, NEW ZEALAND LISTENER (Dec. 18, 2019), https://www.noted.co.nz/tech/tech-tech/cybercrime-most-doesnt-involve-computer-hacking [https://perma.cc/6Q8E-4CS5] (noting that humans are often the weak point in information technology ecosystems).

[20] The current (as of this writing) electoral campaign in Great Britain may well provide a proving ground for this claim. *See* Cat Zakrzewski, *The Technology 202: U.K. Elections Provide Key Test for American Tech Companies' Efforts to Fight Disinformation*, WASH. POST (Dec. 12, 2019), https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2019/12/12/the-technology-202-u-k-elections-provide-key-test-for-americ

This Essay attempts three things. First, it examines and questions assumptions about the 2016 disinformation campaign through the lens of hacking. Second, it looks for other examples—test cases—of successful informational interventions for fun and profit. Finally, it closes with suggestions about what voters, governments, and platforms can do to reduce the likelihood of future successful attacks.

I. THE USUAL SUSPECTS

The standard story about Russian intervention is constructed on a number of assumptions. These might, or might not, be warranted. But, the point of enumerating them is to avoid the mistake of rounding up the usual suspects in response to the commission of a crime. No serious commentator doubts that Russia tried hard to influence the outcome of the 2016 election.[21] Constructing an effective defense, though, requires a careful understanding of how they did so and why it worked.

The first and perhaps most important question is whether Russia's intervention can be successfully replicated.[22] Nearly all observers were surprised by the outcome of the 2016 presidential race.[23] Even the attackers may not know what worked or why. If they do, they will certainly attempt to repeat their success. But an

an-tech-companies-efforts-to-fight-disinformation/5df1283588e0fa51665c097a/ [https://perma.cc/375X-L4U6].

[21] There is no shortage of non-serious commentators who doubt Russian interference. *See, e.g.*, Catie Edmondson, *G.O.P. Senators, Defending Trump, Embrace Debunked Ukraine Theory*, N.Y. TIMES (Dec. 3, 2019), https://www.nytimes.com/2019/12/03/us/politics/republicans-ukraine-conspiracy-theory.html [https://perma.cc/R5CR-DGKZ]; Adam Gabbatt, *Trump Resurfaces Debunked Theory Ukraine Interfered in 2016 Election*, GUARDIAN (Nov. 22, 2019, 15:10), https://www.theguardian.com/us-news/2019/nov/22/donald-trump-resurfaces-debunked-theory-ukraine-interfered-2016-election [https://perma.cc/9AVA-SD43].

[22] A second-order question is how to prioritize information-based reforms. Russia's efforts undoubtedly changed votes. So did the near obsession by mainstream media sources, such as the *New York Times*, with ultimately irrelevant questions about Secretary Clinton's private e-mail server. *See* Erik Wemple, *Studies Agree: Media Gorged on Hillary Clinton Email Coverage*, WASH. POST (Aug. 25, 2017, 3:44 PM), https://www.washingtonpost.com/blogs/erik-wemple/wp/2017/08/25/studies-agree-media-gorged-on-hillary-clinton-email-coverage/ [https://perma.cc/7XE4-ZXS4]. And so, too, did the transformation of channels such as Fox News into de facto agents of the Trump campaign. *See generally* Garrett M. Graff, *Fox News Is Now a Threat to National Security*, WIRED (Dec. 11, 2019, 8:00 AM), https://www.wired.com/story/fox-news-is-now-a-threat-to-national-security/ [https://perma.cc/7XE4-ZXS4]. Foreign informational interventions might be more effective or more objectionable, but advocates of reform should consider that there are limited resources to address information problems and prioritize accordingly.

[23] *See* Hensch, *supra* note 13.

intervention based on, for example, broad exposure and Bayesian learning[24] might provide little guidance to hackers.

The second question is how strong an influence the Russian intervention exerted. The hacking metaphor assumes that the attacks changed votes—in other words, it assumes that, but for Russian efforts, a decisive number of voters would have voted for the Democratic candidate instead of the Republican one. However, the null hypothesis, despite how depressing observers may find it, is that voters preferred Trump to Clinton, even if this meant supporting policies (such as tax cuts and reductions in social safety net programs) that might negatively affect their economic interests.[25] One interesting way of examining this question—if the data are available—would be to look at ticket-splitting in social media users and non-social media users. Controlling for other factors (not an easy task), if social media users more often voted for candidate Trump, but Democratic candidates for other positions than non-users did, that is at least suggestive of an effect from disinformation on social media.[26] Internal inconsistencies in voter preferences could thus be telling.

The disinformation campaign also appears to have had multiple goals, some of which conflict. For example, it is plain that Russia wanted to create political and social discord among Americans, as well as reducing political support for candidate Clinton.[27] A well-known example is the use of parallel disinformation strategies to generate a clash in front of an Islamic center in Houston, Texas. Russian operatives set up two opposing Facebook groups, the "Heart of Texas" (which purported to resist the "Islamization of Texas") and the "United Muslims for America" (which

[24] *See, e.g.*, Jonny Brooks-Bartlett, *Probability Concepts Explained: Bayesian Inference for Parameter Estimation*, TOWARDS DATA SCIENCE (Jan. 5, 2018), https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348 [https://perma.cc/TE82-ARUQ] (providing an explanation of Bayesian inference and its underlying theory). An example of a Bayesian technique to deal with information problems is filtering of spam e-mail messages. *See* Pieter Arntz, *Explained: Bayesian Spam Filtering*, MALWAREBYTES LABS (Feb. 17, 2017), https://blog.malwarebytes.com/security-world/2017/02/explained-bayesian-spam-filtering/ [https://perma.cc/F3VQ-JXEK].

[25] People routinely vote on a heterogeneous set of preferences. A number of studies suggest that non-economic factors predominated for key constituencies. *See, e.g.*, Ann M. Oberhauser, Daniel Krier & Abdi M. Kusow, *Political Moderation and Polarization in the Heartland: Economics, Rurality, and Social Identity in the 2016 U.S. Presidential Election*, 60 SOC. Q. 224 (2019) (analyzing Iowa voters); Tyler T. Reny, Loren Collingwood & Ali A. Valenzuela, *Vote Switching in the 2016 Election: How Racial and Immigration Attitudes, Not Economics, Explain Shift in White Voting*, 83 PUB. OPINION Q. 91 (2019) (providing that racial and immigration concerns affected the voting behavior of White voters).

[26] Ideally, an assessment would include a wider set of variables, such as whether a user saw or "liked" any of the disinformation postings, commented on them, retweeted them, etc.

[27] *See* MUELLER, *supra* note 11, at 4, 14.

claimed to want to "Save Islamic Knowledge").[28] Each Facebook page called upon its followers to engage in a demonstration in front of the Islamic center on May 21, 2016.[29] Only about a dozen protesters supporting the "Heart of Texas" position showed up; the crowd of counterprotesters was far larger.[30] This effort to intensify existing conflicts could be quite successful in heightening tensions and undermining social cohesion. But it also seems likely to encourage voter turnout rather than to suppress it: people motivated enough to show up for an in-person demonstration are likely to be motivated to go to the polls and vote.[31]

A third question is whether micro-targeted disinformation is successful without the cooperation of members (particularly influential ones) of the targeted group.[32] The Twitter account Blacktivist, set up by Russian Internet operatives, pushed for a demonstration in Baltimore on the anniversary of the death of Freddie Gray, who died due to mistreatment while in police custody.[33] Blacktivist reached out to Reverend Heber Brown, III, a Baltimore minister, to try to line him up as a supporter.[34] Brown, though, resisted Blacktivist's efforts, concerned that they originated outside Baltimore and might undermine local community efforts.[35] Social media offers a vector for reaching individuals without using traditional channels or

[28] *See* Todd J. Gilman, *Russian Trolls Orchestrated 2016 Clash at Houston Islamic Center*, *New Senate Intel Report Recalls*, DALL. MORNING NEWS (Oct. 8, 2019, 12:56 PM), https://www.dallasnews.com/news/politics/2019/10/08/russian-trolls-orchestrated-2016-clash-houston-islamic-center-senate-intel-report-says/ [https://perma.cc/WDC3-4NGA].

[29] Claire Allbright, *A Russian Facebook Page Organized a Protest in Texas. A Different Russian Page Launched the Counterprotest.*, TEX. TRIBUNE (Nov. 1, 2017, 4:00 PM), https://www.texastribune.org/2017/11/01/russian-facebook-page-organized-protest-texas-different-russian-page-l/ [https://perma.cc/JB93-ERV5].

[30] *See* Scott Shane, *How Unwitting Americans Encountered Russian Operatives Online*, N.Y. TIMES (Feb. 18, 2018), https://www.nytimes.com/2018/02/18/us/politics/russian-operatives-facebook-twitter.html [https://perma.cc/FLX9-74MQ].

[31] *Cf.* Stephen Coleman, *The Effect of Social Conformity on Collective Voting Behavior*, 12 POLIT. ANALYSIS 76 (2004); Leonie Huddy & Nadia Khatib, *American Patriotism, National Identity, and Political Involvement*, 51 AM. J. POLIT. SCI. 63, 73–74 (2007).

[32] *See generally* Dipayan Ghosh, *Banning Micro-Targeted Political Ads Won't End the Practice*, WIRED (Nov. 22, 2019, 12:20 PM), https://www.wired.com/story/banning-micro-targeted-political-ads-wont-end-the-practice/ [https://perma.cc/32B4-B5RK] (discussing the response of Facebook, Google, and Twitter toward paid political advertising).

[33] *See* Jason Parham, *Russians Posing as Black Activists on Facebook Is More Than Fake News*, WIRED (Oct. 18, 2017, 9:00 AM), https://www.wired.com/story/russian-black-activist-facebook-accounts/ [https://perma.cc/XBC3-FSPG] (discussing how certain groups have infiltrated others on social media with fabricated accounts and targeted ads).

[34] *See* Alison Knezevich & Justin Fenton, *'Blacktivist' Account Linked to Russia Raised Suspicion Among Baltimore Activists*, BALT. SUN (Sept. 29, 2017), https://www.baltimoresun.com/maryland/bs-md-ci-blacktivist-social-media-20170929-story.html [https://perma.cc/3G6R-WVSB].

[35] *Id.*

satisfying traditional gatekeepers. However, disinformation efforts that are not successful in persuading key influencers are much less likely to succeed.

Lastly, the standard narrative on Russian interference in 2016—or, indeed, any disinformation effort—relies on an (often unstated) view of what the desired outcome should have been. This could be a normative claim: disinformation caused American voters to elect a candidate manifestly unsuited for the presidency, or it caused parents to avoid vaccinating their children due to fears about potential side effects.[36] The claims can also be descriptive: positing that disinformation causes people to make political, financial, or personal choices that are misaligned with their ex ante preferences. Both normative and descriptive claims about disinformation's effects are open to challenge. Normatively, a stable minority of American voters (around 40–43%), when polled, strongly support President Trump and his policies.[37] It is possible, of course, that the disinformation is sticky[38]—it changed voters' views before the election, and those views have remained constant.[39] That leads into the second challenge: it is not clear why people's ex ante views create the correct baseline for analysis. Those views themselves may have been the product of previous informational interventions, and humans generally do not have reliable access to their preference ordering (or, perhaps, may not have reliable preference orderings).

These assumptions about Russia's disinformation intervention may be correct. It may be difficult to test or disprove them. But they are worth surfacing as the United States tries to prevent future interference in its affairs based on inaccurate online information campaigns.[40]

[36] *See, e.g.*, Lesley Chiou & Catherine Tucker, Fake News and Advertising on Social Media: A Study of the Anti-Vaccination Movement (July 27, 2018) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3209929 [https://perma.cc/FQ36-BM7M] (studying "the role of social networks and advertising on social networks in the dissemination of false news stories about childhood vaccines").

[37] *See How Unpopular Is Donald Trump?*, FIVETHIRTYEIGHT, https://projects.fivethirtyeight.com/trump-approval-ratings/ [https://perma.cc/DFR6-KWVF] (last visited Mar. 17, 2020) (providing "an updating calculation of the president's approval rating").

[38] *See, e.g.*, Drew Harwell, *Doctored Images Have Become a Fact of Life for Political Campaigns. When They're Disproved, Believers 'Just Don't Care.*,*'* WASH. POST (Jan. 14, 2020, 5:00 AM), https://www.washingtonpost.com/technology/2020/01/14/doctored-political-images/ [https://perma.cc/V3RD-ZT2R] (discussing the impact of technology and social media on the practice of sharing doctored images of political rivals).

[39] This is plausible: people often evince confirmation bias, where they seek out information that reinforces rather than challenges their existing beliefs.

[40] And, as always, there is the risk that reform efforts will be deployed strategically, against domestic political opposition, as well as, or instead of, being used against foreign interference. *See, e.g.*, Kirsten Han, *Want to Criticize Singapore? Expect a 'Correction Notice*,*'* N.Y. TIMES (Jan. 21, 2020), https://www.nytimes.com/2020/01/21/opinion/fake-news-law-singapore.html [https://perma.cc/U74A-PG9U]. *See generally* Derek E. Bambauer, *Against Jawboning*, 100 MINN. L. REV. 51 (2015) (discussing attempts by several

II. SUCCESSFUL HACKS

Information hacking is neither new nor exclusively human. Animals misrepresent information for a variety of reasons.[41] Edible butterflies evolve to resemble noxious neighbors.[42] Fireflies from one species imitate mating signals from another, then eat their suitors.[43] Cuttlefish change color to improve their odds of mating.[44] Some disinformation mechanisms in animals are involuntary: butterflies that are poor mimics due to genetic chance or development are more likely to become prey, and thus less likely to pass on genes.[45] Evolution is a harsh judge. But some tactics, like those of the cuttlefish, are the result of cognition, and thus fall closer to human disinformation strategies.[46]

Humans propagate deliberate false information for a variety of reasons as well: to amuse, to profit, to obtain power, and to irritate.[47] Some examples of successful disinformation produce results through a process analogous to evolution: large-scale variation over time with a feedback loop sorts winners from losers. For example, scammers send out huge numbers of unsolicited e-mail messages promoting a variety of penny stocks. Gullible investors buy the stocks—not in great numbers, but enough to make the scheme worthwhile.[48] The low cost of penny stock tout spam means that it does not need to be particularly well-targeted; one study found that

state attorneys general to pressure Google into dealing with copyright infringement and the content showing up in search results).

[41] For a readable set of examples, see Barbara J. King, *Deception in the Wild*, 321 SCI. AM. 50, 52–53 (2019).

[42] *See, e.g.*, Mitsuho Katoh, Haruki Tatsuta & Kazuki Tsuji, *Rapid Evolution of a Batesian Mimicry Trait in a Butterfly Responding to Arrival of a New Model*, 7 SCI. REP. 6369, 6369–40 (2017), https://doi.org/10.1038/s41598-017-06376-9 [https://perma.cc/JX2B-Q8TL].

[43] *See, e.g.*, James E. Lloyd, *Aggressive Mimicry in Photuris Fireflies: Signal Repertoires by Femmes Fatales*, 187 SCI. 452, 452–53 (1975).

[44] *See, e.g.*, Culum Brown et al., *It Pays to Cheat: Tactical Deception in a Cephalopod Social Signaling System*, 8 BIO. LETTERS 729, 730 (2012).

[45] *See* Katoh, Tatsuta & Tsuji, *supra* note 42, at 6369–40.

[46] *See* King, *supra* note 41, at 54 (also citing example of canines).

[47] *See generally* Mark Verstraete, Derek E. Bambauer & Jane R. Bambauer, *Identifying and Countering Fake News* (Ariz. Legal Studies, Discussion Paper No. 17-15, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3007971 [https://perma.cc/5CWT-6Y33] (classifying different types of fake news).

[48] *See* Karen K. Nelson, Richard A. Price & Brian R. Rountree, *Are Individual Investors Influenced by the Optimism and Credibility of Stock Spam Recommendations?*, 40 J. BUS. FIN. & ACCT. 1155, 1158–61 (2013).

spammers can earn over 4% return on this tactic[49] (before transaction costs, which are low[50]).

There are other, more subtle ways to profit from disinformation—ones that emphasize the sheer difficulty of categorizing data and motives in this space.[51] For example, financial analyst Harry Markopoulos, who famously outed the criminal conduct by Bernie Madoff,[52] issued a recent report stating that General Electric ("GE") was "a bigger fraud than Enron."[53] Markopoulos's claim rests on a plausible, but unlikely, theory that GE has massive undocumented liabilities on its accounting books from its long-term care insurance business.[54] Markopoulos, though, secretly shared his conclusions with an unnamed hedge fund that shorted GE's stock ahead of its release, after shopping the information to three other funds.[55] (The stock fell by 11% on the day the report debuted.[56]) In exchange, Markopoulos received a share of the fund's profits from those trades.[57] GE has been a case study in mismanagement for years, but its accounting is viewed as rigorous by many outside accounting experts.[58] So, it is possible that Markopoulos is a whistleblower, out to save investors from fraud. It is more likely that Markopoulos is trading on his Madoff fame to peddle a shoddy theory that makes the market rather than correcting it.[59]

[49] *See* Laura Frieder & Jonathan Zittrain, *Spam Works: Evidence from Stock Touts and Corresponding Market Activity*, 30 HASTINGS COMM. & ENT. L.J. 479, 501 (2008).

[50] *See, e.g.*, Derek E. Bambauer, *Solving the Inbox Paradox: An Information-Based Policy Approach to Unsolicited E-mail Advertising*, 10 VA. J.L. & TECH. 1, 11–14 (2005).

[51] *See* Verstraete, Bambauer & Bambauer, *supra* note 47, at 8–9.

[52] *See* Shawn Tully, *How the Man Who Nailed Madoff Got GE Wrong*, FORTUNE (Oct. 3, 2019, 4:00 AM), https://fortune.com/2019/10/03/ge-accounting-markopolos-madoff/ [https://perma.cc/T38K-HUW5].

[53] *See* HARRY MARKOPOULOS, GENERAL ELECTRIC, A BIGGER FRAUD THAN ENRON, https://fm.cnbc.com/applications/cnbc.com/resources/editorialfiles/2019/8/15/2019_08_15_GE_Whistleblower_Report.pdf [https://perma.cc/Z8CP-LQ94] (last visited Mar. 5, 2020).

[54] *See* Tully, *supra* note 52.

[55] *See* Mark Vandevelde, *Harry Markopolos: The Scourge of Madoff Trains His Sights on GE*, FIN. TIMES (Aug. 16, 2019), https://www.ft.com/content/243c4728-c00c-11e9-b350-db00d509634e [https://perma.cc/FE63-DZBE].

[56] *See* Tully, *supra* note 52.

[57] *Id.*

[58] *Id.*

[59] Crediting Markopoulos with unmasking Bernie Madoff may also be inaccurate information, likely due to hindsight bias. For any investor or firm of any size, there will be a Cassandra or two who claim that the operation is a sham. This is also what makes Michael Lewis's book THE BIG SHORT an analytical disappointment: there is insufficient evidence that his protagonists were insightful, rather than lucky, in predicting the weakness in the home mortgage market and collateralized debt obligations. *See generally* MICHAEL LEWIS, THE BIG SHORT (2010); *but cf. generally* JOHN ALLEN PAULOS, A MATHEMATICIAN PLAYS THE STOCK MARKET (2003) (explaining how popular investment approaches and theories are incorrect and unable to help an investor, even a professional mathematician, make sense of

People also lie to increase their chances of mating.[60] Men misrepresent themselves as taller than they actually are.[61] Women claim to be younger.[62] Everyone lies about their income and uses photos of their younger selves as profile pictures.[63] Disinformation works: a study by the online dating site OkCupid showed that reported income was positively correlated with the number of messages a user received, especially for men.[64] It may be difficult to detect inaccurate information—requests for a bank statement on a first date are likely to be received poorly—or liars may be exploiting transaction costs. Luring a potential partner into communication, or an in-person meeting, gives the deceiver a chance to make a good impression on other grounds or to perpetuate the fiction.[65]

Sometimes, the goal of disinformation is expressive, not pecuniary: to advance a particular normative view or to undermine one. The online movie review site Rotten Tomatoes changed its crowdsourced ratings feature to prevent fake pre-release reviews that targeted movies such as *Captain Marvel*, *Star Wars: The Last Jedi*, and *Black Panther*.[66] Online critics—trolls, more accurately—attacked these

the stock market and accurately predict its movements). Put colloquially, every year amateur basketball fans place bets (many illegal) on a sixteenth-seeded team to upset a first-seeded team in the NCAA Men's Basketball Championship. Such an upset has occurred once, when the University of Virginia lost to the University of Maryland – Baltimore County in 2018. The governor of Maryland, Larry Hogan, had UMBC defeating Virginia in his bracket. *See* Michelle R. Martinelli, *Maryland Governor, Senator Predicted UMBC's Upset and Picked Retrievers as National Champions*, USA TODAY (Mar. 17, 2018, 12:51 AM), https://ftw.usatoday.com/2018/03/umbc-maryland-baltimore-county-upset-uva-virginia-ncaa-tournament-march-madness-prediction-larry-hogan-chris-van-hollen [https://perma.cc/7RAS-NS4G] (showing that Gov. Hogan was right but would be an unlikely source of future wagering advice for NCAA bracket lovers).

[60] *See generally* Irina D. Manta, *Tinder Lies*, 54 WAKE FOREST L. REV. 207 (2019).

[61] *See The Big Lies People Tell In Online Dating*, OKCUPID (July 7, 2010), https://theblog.okcupid.com/the-big-lies-people-tell-in-online-dating-a9e3990d6ae2 [https://perma.cc/42ZM-4CTE].

[62] *Id.*

[63] *Id.*

[64] *Id.*

[65] *See* Brian Fung, *OkCupid Reveals It's Been Lying to Some of Its Users. Just to See What'll Happen.*, WASH. POST (July 28, 2014, 11:57 AM MDT), https://www.washingtonpost.com/news/the-switch/wp/2014/07/28/okcupid-reveals-its-been-lying-to-some-of-its-users-just-to-see-whatll-happen/ [https://perma.cc/9HEC-ADUH] (describing experiments OkCupid conducted on its users that showed users were more likely to respond to messages and exchange contact information with another user when OkCupid removed the profile photo of the user making the initial contact).

[66] George Nash, *Rotten Tomatoes Rescues* Captain Marvel *from Review Trolls*, GUARDIAN (Feb. 27, 2019, 11:30 AM), https://www.theguardian.com/film/2019/feb/27/rotten-tomatoes-captain-marvel-brie-larson-review-trolls [https://perma.cc/G5FX-BH73]; *see also* Alex Abad-Santos, *How* Captain Marvel *and Brie Larson Beat the Internet's Sexist Trolls*, VOX (Mar. 11, 2019, 11:10 AM), https://www.vox.com/culture/2019/3/8/18254584/captain-marvel-boycott-controversy [https://perma.cc/YK6A-TBZP].

films for purportedly advancing a political agenda by casting women and people of color in lead roles.[67] The short-run mechanism was financial; if the movies tanked at the box office, perhaps Hollywood's studios would return to reserving key positions for white men.[68] But the trolls would receive no monetary compensation even if they succeeded. Instead, their objective was less tangible: to resist calls for more diversity both in motion pictures and in the journalists who cover them. Aided by the Rotten Tomatoes change and a tweak to YouTube's algorithm,[69] *Captain Marvel* triumphed over the trolls, earning over $153 million in its opening weekend and over $1.1 billion in revenues worldwide during its release.[70] But films such as the all-female *Ghostbusters* remake suffered from trolling,[71] and similar online complaints appear to have led Disney to jettison the initial script for the final *Star Wars* movie, *The Rise of Skywalker*.[72] Disinformation may thus use money as a lever rather than as a goal.

[67] *See* Abad-Santos, *supra* note 66.

[68] *See* Yohana Desta, *Rotten Tomatoes Is Fighting Back Against White Nationalist* Black Panther *Trolls*, VANITY FAIR (Feb. 2, 2018), https://www.vanityfair.com/hollywood/2018/02/rotten-tomatoes-black-panther-facebook-group [https://perma.cc/K54S-GRH6] (reporting on organized efforts by white nationalists to write trolling film reviews on Rotten Tomatoes).

[69] James Hale, *Here's How YouTube Fought* 'Captain Marvel' *Trolls,* TUBEFILTER (Mar. 8, 2019), https://www.tubefilter.com/2019/03/08/heres-how-youtube-fought-captain-marvel-trolls/ [https://perma.cc/Z66V-MYP3].

[70] *Captain Marvel (2019)*, THE NUMBERS, https://www.the-numbers.com/movie/Captain-Marvel-(2019)#tab=summary [https://perma.cc/MMK5-Z9DC].

[71] *See* Emma Grey Ellis, *Trolls Are Boring Now*, WIRED (Mar. 13, 2019, 11:22 AM), https://www.wired.com/story/trolls-are-boring/ [https://perma.cc/7LHU-UENL] (explaining that the tactics that failed against *Captain Marvel* worked with great effect to damage the box office success of the new, all-female version of *Ghostbusters*).

[72] *See* Alex Abad-Santos & Alissa Wilkinson, Star Wars: The Rise of Skywalker *Was Designed to Be the Opposite of* The Last Jedi, VOX (Dec. 27, 2019, 12:20 PM), https://www.vox.com/culture/2019/12/27/21034725/star-wars-the-rise-of-skywalker-last-jedi-j-j-abrams-rian-johnson [https://perma.cc/75C3-HWFS] (discussing the online backlash against the trilogy's previous installment, *The Last Jedi*, and explaining how *The Rise of Skywalker* seems to reflect that Disney was aware of the criticism and made changes in response) ; Britt Hayes, *Turns Out Colin Trevorrow's Version of* Star Wars: Episode IX *Was Good, Actually*, AV CLUB (Jan. 14, 2020, 2:32 PM), https://news.avclub.com/turns-out-colin-trevorrows-version-of-star-wars-episod-1841002112?fbclid=IwAR0kdfhnxbvXBBJs HkOu6MNhpVytk_ZkUw-9NhojqZW5-f0th9UqibkXCi0 [https://perma.cc/4B45-KFXP] (discussing the plot points of the original script for Episode Nine of the *Star Wars* film franchise that was rejected when J.J. Abrams was brought back as the director and co-writer of *The Rise of Skywalker*); Brian Lowry, 'The Rise of Skywalker' *Takes Flight After the Rise of the* 'Star Wars' *Trolls*, CNN (Dec. 13, 2019, 12:50 PM), https://www.cnn.com/2019/12/13/entertainment/star-wars-trolls-trnd/index.html [https://perma.cc/6HQ7-KTK8] (noting the rise in vitriolic trolling regarding the *Star Wars* franchise, including "evidence of deliberate, organized political influence measures disguised as fan arguments").

The prevalence of prevarication strategies may seem depressing.[73] However, these case studies offer a set of testbeds for potential interventions, as well as enabling after-action analysis of efforts that did not work.

III. HACKING BACK

Cybersecurity offers a number of lessons for how to respond to information hacking.[74] Unfortunately, some of these lessons concern the limits of preventing and remediating hacks. But dispelling false hope is a service in itself. These insights can be summarized in three key points. First, preventing information hacks is difficult if not impossible. Second, remediating hacks shows more promise, but recovery tends to be unpleasant and expensive, particularly for political issues. Finally, while educating users is a perennially popular solution, it is an illusory one that may actually be counterproductive.

A. The Limits of Prevention

Cybersecurity demonstrates the profound challenges of preventing successful attacks, which should induce skepticism about most current approaches to disinformation.[75] Blocking attacks is a popular strategy for computer security. It

[73] *See generally* Dawn Carla Nunziato, *The Marketplace of Ideas Online*, 94 NOTRE DAME L. REV. 1519 (2019) (describing how bad actors continue to interfere with the marketplace(s) of ideas).

[74] *See generally* Ellen Nakashima, *U.S. Cybercom Contemplates Information Warfare to Counter Russian Interference in 2020 Election*, WASH. POST (Dec. 25, 2019, 3:45 PM), https://www.washingtonpost.com/national-security/us-cybercom-contemplates-information-warfare-to-counter-russian-interference-in-the-2020-election/2019/12/25/21bb246e-20e8-11ea-bed5-880264cc91a9_story.html [https://perma.cc/5QV8-5KVZ].

[75] Most of the political discourse about disinformation on social media platforms results in pressure on these firms to purge their sites of deliberately inaccurate data. *See, e.g.*, Brian Fung, *Facebook to Ban Census Suppression on Its Platforms,* CNN (Dec. 19, 2019, 11:37 AM), https://www.cnn.com/2019/12/19/tech/facebook-census-suppression-policy/index.html [https://perma.cc/7YPA-JAVU]; Oliver Milman, *Defiant Mark Zuckerberg Defends Facebook Policy to Allow False Ads*, GUARDIAN (Dec. 2, 2019, 9:19 AM), https://www.theguardian.com/technology/2019/dec/02/mark-zuckerberg-facebook-policy-fake-ads [https://perma.cc/554S-VER5]. Most scholarly reform proposals also concentrate on blocking or removing suspect information, often by imposing liability on platforms. *See generally* Annemarie Bridy, *Remediating Social Media: A Layer-Conscious Approach,* 24 B.U. J. SCI. & TECH. L. 193, (2018) (providing a high-level regulatory history of online speech and arguing that "adopting a must-carry obligation for social media platforms is not what the Internet needs"); Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401 (2017) (arguing that statutory protections for websites are overly broad under the Communications Decency Act); Richard L. Hasen, *Cheap Speech and What It Has Done (To American Democracy)*, 16 FIRST AMEND. L. REV. 200 (2017) (analyzing the costs associated with the rise of speech

explains why users and organizations are admonished by security experts to patch their systems, run intrusion detection and anti-virus software, filter e-mail messages for phishing attempts, share information about threats, and inculcate users with caution about all things Internet.[76] Yet systems are continually compromised due to misconfiguration, complexity, and simple human gullibility.[77] Prevention is conceptually simple but difficult in implementation.

Information problems are even harder to prevent. Patching humans is much more difficult than patching software. Behavioral economics has identified a number of cognitive human biases and traits that can be exploited by attackers.[78] However, even when we know, for instance, that hindsight bias is in play, it is hard to ameliorate.[79] These neurological characteristics are susceptible to hacks precisely because they are persistent. This is one of the challenges of efforts to educate people, or to shift social norms in other ways, in response to attacks. Education operates on the implicit assumption that human behavior can be altered in a particular direction with sufficient effort and focus.[80] Cybersecurity research demonstrates the faulty

on the internet and social media and its effects on American democracy); Alexander Tsesis, *Social Media Accountability for Terrorist Propaganda*, 86 FORDHAM L. REV. 605, 610–11 (2017) (arguing that criminal liability "would be the most effective means of addressing the dissemination of extremist digital communications"); Cass R. Sunstein, Falsehoods and the First Amendment (Preliminary draft 7/25/19, July 29, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3426765 [https://perma.cc/3W3Z-SPVU] (arguing that the government should have authority to control fake content and that social media platforms should do more to control the same).

[76] *See, e.g.*, U.S. DEP'T OF HEALTH & HUMAN SERVS., DHHS OFFICE FOR CIVIL RIGHTS, HIPAA SECURITY RULE CROSSWALK TO NIST CYBERSECURITY FRAMEWORK (Feb. 22, 2016), https://www.hhs.gov/sites/default/files/nist-csf-to-hipaa-security-rule-crosswalk-02-22-2016-final.pdf [https://perma.cc/L8E5-DHU3].

[77] *See generally* Mark Evans et al., *Human Behavior as an Aspect of Cybersecurity Assurance*, 9 SECURITY COMM. NETWORKS 4667 (2016) (proposing a framework for cybersecurity assurance); *see also* Micke Ahola, *The Role of Human Error in Successful Cyber Security Breaches*, USECURE (Oct. 18, 2019, 9:46 AM), https://blog.getusecure.com/post/the-role-of-human-error-in-successful-cyber-security-breaches [https://perma.cc/33EQ-3DBX]; Martin Kaste, *Cybercrime Booms as Scammers Hack Human Nature to Steal Billions*, NPR (Nov. 18, 2019, 5:32 AM), https://www.npr.org/2019/11/18/778894491/cybercrime-booms-as-scammers-hack-human-nature-to-steal-billions [https://perma.cc/C8YN-WPSL].

[78] *See, e.g.*, Victoria Fineberg, *BECO: Behavioral Economics of Cyberspace Operations*, 2 J. CYBER SEC. & INFO. SYS. 20 (2016).

[79] *See, e.g.*, Neal J. Roese & Kathleen D. Vohs, *Hindsight Bias*, 7 PERSP. ON PSYCH. SCI. 411, 417–19 (2012).

[80] There may be some hopeful examples. *See, e.g.*, Eliza Mackintosh, *Finland Is Winning the War on Fake News. What It's Learned May Be Crucial to Western Democracy*, CNN (May 2019), https://www.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/ [https://perma.cc/4C8F-BQ6C]. There are some nascent public-sector efforts in the U.S., although it is not clear how they will be implemented. Yael Grauer, *Arizona Now Has a Task*

optimism at play in educational efforts.[81] Security experts have been trying to get computer users to change how they interact online for years, with little success: people still click suspicious links in e-mail messages, open attachments, reuse passwords across websites, connect freely to public wireless networks that lack encryption, and generally engage in other risky behaviors despite ongoing informational campaigns.[82]

Moreover, even limited success may pay major dividends for attackers. Compromising a single computer or user account in an organization can enable malicious actors to expand their control rapidly and widely.[83] Similarly, in a close election, disinformation efforts that affect only a small fraction of voters could prove decisive. The heterogeneous preferences of voters present a wide array of weakness that information hacking can exploit. This makes defensive efforts much more difficult. Attackers always have the advantages of time, numbers, and the initiative. They are highly motivated—by the prospect of political gain, financial benefit, or both. Some hacks require substantial expertise to execute successfully. But with both information and cybersecurity attacks, some can be packaged or automated in a way that allows less-skilled actors to mount successful interventions.[84]

In short, determined attackers enjoy significant advantages in hacking humans and computers alike. Preventing exploits is an appealing but unrealistic option.[85]

Force Focused on Countering Disinformation, SLATE (Dec. 18, 2019, 7:30 AM), https://slate.com/technology/2019/12/arizona-task-force-disinformation-judicial-system.html [https://perma.cc/Q27L-Z7J4].

[81] *See* Derek E. Bambauer, *Ghost in the Network*, 162 U. PA. L. REV. 1011, 1043–47 (2014).

[82] *See, e.g.*, Michael Greene, *The Password Reuse Problem Is a Ticking Time Bomb*, HELP NET SECURITY (Nov. 12, 2019), https://www.helpnetsecurity.com/2019/11/12/password-reuse-problem/ [https://perma.cc/9AYU-EWP3]; Kaste, *supra* note 77; CowBear Bebop, *How to Secure Your Customers' Data Over Insecure Public Wi-Fi*, TUNNELBEAR BLOG (Sept. 23, 2019), https://www.tunnelbear.com/blog/how-to-secure-your-customers-data-over-insecure-public-wi-fi/ [https://perma.cc/M8YG-5N7M]; *Successful White House Spear Phishing Attacks Show No One Is Safe*, GRAPHUS BLOG (Jan. 21, 2020), https://www.graphus.ai/successful-white-house-spear-phishing-attacks-show-no-one-is-safe/ [https://perma.cc/XR4V-Q79M].

[83] *See, e.g.*, Michael Kassner, *Anatomy of the Target Data Breach: Missed Opportunities and Lessons Learned*, ZDNET (Feb. 2, 2015, 4:29 PM), https://www.zdnet.com/article/anatomy-of-the-target-data-breach-missed-opportunities-and-lessons-learned/ [https://perma.cc/9KQM-WUMP].

[84] *See, e.g.*, Peter P. Swire, *A Theory of Disclosure for Security and Competitive Reasons: Open Source, Proprietary Software, and Government Systems*, 42 HOUS. L. REV. 1333, 1350 n.43 (2006) (discussing "script kiddies").

[85] *See*, Bambauer, *supra* note 81, at 1019–20 (arguing that preventing exploitation is inevitably futile).

B. The Costs of Remediation

A more promising alternative to preventing attacks is to undertake measures that reduce their effectiveness or mitigate their impacts. These range from the simple—backing up data and practicing recovery procedures—to the complex—using techniques such as air gaps between key networks and the public Internet,[86] or ensuring that an organization uses heterogeneous operating systems and applications to prevent a single exploit from devastating a monoculture computing environment. These methods are often effective, but they are also expensive, sometimes slow, and raise hard questions about determining what data can be treated as accurate. For example, the city of Baltimore, Maryland has already spent over $18 million to reconstruct its information systems after a pair of ransomware attacks crippled its public services infrastructure.[87]

Disinformation can also be addressed after the fact. Intermediaries such as journalistic organizations can opine on the accuracy of information on-line, and platforms can mark it as suspect, redirect users to other sources, or take the data down. Voters can cease supporting a candidate or cause, and in some instances may be able to act politically, such as via recall efforts, impeachment, or ballot referenda.

A major challenge for remediation of information hacking is that the time window for doing so is often limited. Voters who learn that they have relied upon disinformation in making their selections at the ballot box have little if any hope of changing their decisions. America does not reboot elections. Similarly, most vaccines must be administered within a given period of time, especially for children. Once that period passes, the vaccine will have less efficacy, if any. Policy decisions such as addressing climate change have longer time frames for intervention, but here too, scientists warn that some global effects are already irreversible on any meaningful scale and that others will soon become so.

There are also familiar obstacles from cybersecurity to mitigation tactics. One method is to identify sources of suspect information or to block them altogether. Attribution, however, is a long-standing challenge online. Users can migrate accounts, use automated methods such as bots to disseminate information, or rely on fellow travelers to post their messages.[88] And attackers may well pose as defenders, reporting other users (including accurate sources of information) as suspect both to impede their efforts and to disguise their own malfeasance.[89] These problems with identifying disinformation via source rather than content also affect

[86] *See* Zetter, *supra* note 14 (describing how Stuxnet bridged the air gap between Iran's nuclear centrifuges and the publicly-connected Internet).

[87] *See* Bruce Sussman, *Baltimore, $18 Million Later: 'This Is Why We Didn't Pay the Ransom,'* SECUREWORLD (June 12, 2019, 7:30 AM), https://www.secureworldexpo.com/industry-news/baltimore-ransomware-attack-2019 [https://perma.cc/J3LY-GZ67].

[88] *See* Derek E. Bambauer, *Conundrum*, 96 MINN. L. REV. 584, 589–90, 595–98 (2011).

[89] *See* MARTIN C. LIBICKI ET AL., RAND, THE DEFENDER'S DILEMMA: CHARTING A COURSE TOWARD CYBERSECURITY 26 (2015), https://www.rand.org/content/dam/rand/pubs/research_reports/RR1000/RR1024/RAND_RR1024.pdf [https://perma.cc/W5ZG-82FB].

the prospect of information sharing to address disinformation. A given source may have different names on different sites, may use tactics such as Virtual Private Networks to conceal their origin, and may attempt to set up their own third-party verifiers or certifiers (a process known as astroturfing or greenwashing).[90]

There are similar challenges for information sharing efforts about suspect content. This type of collaboration has been popular in American cybersecurity efforts in both the public and private sectors.[91] A panoply of intermediaries has been created to distribute information about vulnerabilities, threats, exploits, and safeguards.[92] However, organizations on different systems may find this data to be of limited value—if a firm's web server runs the Linux operating system, data about Windows bugs will not help much. And even when an entity has timely, accurate data about a threat, it may lack the resources to mitigate it. Purchasing new systems takes time and resources, as does patching existing ones. Smaller organizations may not have the necessary expertise to react quickly or may not have an accurate picture of what comprises their information technology environment.

Thus, a combination of legal and practical constraints cabins the potential for remediation or mitigation techniques to address information hacking.

C. The Illusory Promise of Education

One well-worn approach to cybersecurity issues is education.[93] By training users to be more sophisticated and skeptical about their interactions with information technology, attacks such as spearphishing[94] and password guessing will become less likely to succeed. This approach is popular because it is seemingly straightforward and not particularly costly.[95] In addition, it serves a valuable function in the political economy of apportioning responsibility for disinformation problems and fixes. However, as appealing as education is, it is unlikely to succeed. Moreover, educational efforts about disinformation may, ironically, serve the second-order goals of attackers by weakening people's perceptions of the competence and efficacy of key societal institutions.[96]

[90] *See* Derek E. Bambauer, *Cybersieves*, 59 DUKE L.J. 377, 439–41 (2009).

[91] *See* Derek E. Bambauer, *Sharing Shortcomings*, 47 LOY. U. CHI. L.J. 465, 465–67 (2015).

[92] *Id.* at 469.

[93] *See, e.g.*, ELIZABETH BODINE-BARON ET AL., COUNTERING RUSSIAN SOCIAL MEDIA INFLUENCE, RAND 46–50 (2018), https://www.rand.org/content/dam/rand/pubs/research_reports/RR2700/RR2740/RAND_RR2740.pdf [https://perma.cc/GB43-BKRA].

[94] *See* GRAPHUS BLOG, *supra* note 82.

[95] *See* Bambauer, *supra* note 81, at 1043–45.

[96] *See* Mark Verstraete & Derek E. Bambauer, *Ecosystem of Distrust*, 16 FIRST AMEND. L. REV. 129, 142 (2018).

Educational efforts are nothing new; calls for greater media literacy have a long history.[97] Online information sharpens the problem since authors and distributors can evade identification and attribution with greater ease than in offline contexts.[98] And, with the rise of communication channels such as e-mail and social media platforms, attackers can increasingly tailor informational efforts to increase their perceived credibility, to leverage human cognitive biases, and to appear as though they originate from a trusted source.[99] Security education has sought to teach users to avoid attachments sent via e-mail or other mediums from an unknown source.[100] But if they appear to come from a known and trusted source, users are more likely to lower their guard. For example, Jeff Bezos of Amazon had his phone hacked when he received malware over the WhatsApp application from the crown prince of Saudi Arabia.[101] And the computer security firm RSA was hacked when an employee opened what appeared to be a spreadsheet from inside the company that announced their bonus; the file was, in fact, a virus that enabled attackers to exfiltrate sensitive data.[102] Wariness about unknown sources may be achievable, but it is difficult to convince people to maintain skepticism when information seems to originate from a trusted source. Thus, creating more sophisticated and literate online information consumers is a laudable goal but one that is quite hard to achieve.

Emphasizing education to combat disinformation may have at least two undesirable consequences. First, it alters the political economy of interventions by shifting focus from creators or distributors of inaccurate information to consumers of it. Users as a whole are an amorphous group and lack an organized entity to

[97] These are increasingly translated to the online context. *See, e.g.*, Natascha A. Karlova & Karen E. Fisher, *A Social Diffusion Model of Misinformation and Disinformation for Understanding Human Information Behaviour*, 18 INFO. RES. 573 (2013), http://www.informationr.net/ir/18-1/paper573.html#.XpUka8hKjZs [https://perma.cc/9V6T-HKEQ].

[98] *See generally* JANNA ANDERSON & LEE RAINIE, *Theme 5: Tech Can't Win the Battle. The Public Must Fund and Support the Production of Objective, Accurate Information. It Must Also Elevate Information Literacy to Be a Primary Goal of Education*, PEW RES. CTR.: THE FUTURE OF TRUTH AND MISINFORMATION ONLINE (Oct. 19, 2017), https://www.pewresearch.org/internet/2017/10/19/theme-5-tech-cant-win-the-battle-the-public-must-fund-and-support-the-production-of-objective-accurate-information-it-must-also-elevate-information-literacy-to-be-a-primary-goal-of-educat/ [https://perma.cc/24AC-QYB7] (including comments on the internet and security issues dealing with authors, bots, and trolls distributing misinformation in a way not easily negated by current solutions).

[99] *See, e.g.*, Kim Zetter, *Researchers Uncover RSA Phishing Attack, Hiding in Plain Sight*, WIRED (Aug. 26, 2011), https://www.wired.com/2011/08/how-rsa-got-hacked/ [https://perma.cc/MW56-3DNW] [hereinafter Zetter, *RSA Phishing Attack*].

[100] *See* GRAPHUS BLOG, *supra* note 82.

[101] *See* Stephanie Kirchgaessner, *Jeff Bezos Hack: Amazon Boss's Phone 'Hacked by Saudi Crown Prince*,*'* GUARDIAN (Jan. 22, 2020), https://www.theguardian.com/technology/2020/jan/21/amazon-boss-jeff-bezoss-phone-hacked-by-saudi-crown-prince [https://perma.cc/XH67-E9QJ].

[102] *See* Zetter, *RSA Phishing Attack*, *supra* note 99.

represent their interests in these debates. Educational efforts implicitly shift the blame to users for undesirable outcomes and can thus reduce pressure for measures that treat other players in the information ecosystem. Second, if educational efforts succeed, they are likely to do so by increasing skepticism about information sources, even ones that are familiar. This runs the risk of continuing a decades-long trend of decreasing American faith in social, political, and journalistic institutions.[103] Such cynicism is, overtly, a goal of the Russian disinformation campaign.[104] Educational efforts could win the battle but lose the war.

CONCLUSION

Cybersecurity offers a useful framework and some helpful cautionary tales for efforts to address disinformation online. It also suggests a few models for interventions that may help.[105] This section describes quarantines, critical (human) infrastructure, and whitelists as potential interventions that hold promise.

One possibility is for Internet platforms to increase their curation of information, including in some cases by quarantining it. Google and YouTube provide a model for this behavior. Historically, Google has been reluctant to alter the output from its organic search based on its PageRank algorithm.[106] However, the search firm has reacted to problematic search entries in at least two ways. First, Google de-lists material that it has categorized as consisting of child sexual abuse images, terrorist content, and privacy-invading disclosures.[107] Those websites remain available, but their removal from search results makes them harder to find as a practical matter. Second, and more interestingly, Google itself intervenes

[103] *See* Verstraete & Bambauer, *supra* note 96, at 142.

[104] *See, e.g.*, *Fighting Russian Disinformation*, FOREIGN POL'Y (Sept. 30, 2019), https://foreignpolicy.com/podcasts/and-now-the-hard-part/fighting-russian-disinformation/ [https://perma.cc/K3DF-EFHW]; *see also* BODINE-BARON ET AL., *supra* note 93, at 8.

[105] This section describes the interventions based on their potential efficacy. It does not address whether they are politically viable, which is a key criterion. *See* Bambauer, *supra* note 81, at 1037–38.

[106] *See* David Segal, *The Dirty Little Secrets of Search*, N.Y. TIMES (Feb. 12, 2011), https://www.nytimes.com/2011/02/13/business/13search.html [https://perma.cc/UPN5-YNWW]; *see also* Vanessa Fox, *New York Times Exposes J.C. Penney Link Scheme That Causes Plummeting Rankings in Google*, SEARCH ENGINE LAND (Feb. 12, 2011, 6:29 PM), https://searchengineland.com/new-york-times-exposes-j-c-penney-link-scheme-that-causes-plummeting-rankings-in-google-64529 [https://perma.cc/T8HK-P89J].

[107] Google also takes down content when a third party alleges that it violates their copyright and decreases the prominence of material that tries to game the company's algorithm via search engine optimization. *See* Segal, *supra* note 106; *see also* Jennifer M. Urban & Laura Quilter, *Efficient Process or "Chilling Effects"? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 SANTA CLARA HIGH TECH. L.J. 621, 626 (2006) (providing analysis of notices of Google's notice-and-takedown procedure used to gain safe harbor protection under the Digital Millennium Copyright Act).

occasionally when disinformation or other suspect data is highly ranked.[108] For a period of time, the top search result for queries about the term "Jew" was a white supremacist organization.[109] Google did not remove the result but instead added a warning—a text box next to the result explaining what it was and why it had risen to the top of the firm's results.[110] Platforms are already under pressure to adopt the first tactic by removing problematic posts and links, but the rapid mutation of Internet information makes it difficult for social media sites to keep up. The second tactic has not been widely utilized thus far, but it holds at least some promise. In its strong form, platforms could replace known disinformation with accurate information on the same topic. In a milder version, sites could add context and disclaimers when users post problematic links or share disinformation. These moves are fraught for platforms because they must make express value judgments and thereby invite criticism. However, even if sites stuck to relatively obvious disinformation (for example, posts that mention the terms "Obama," "birth certificate," and "Kenya," or ones that reference Pizzagate),[111] that would represent significant progress.

A second effort flows from network theory.[112] Either online or offline, key influencers could be trained to spot disinformation and to intervene. This is a form of targeted user education that is focused; it effectively deputizes certain people in the hunt for bad data. It will require some work to identify who the critical nodes are in a given social network. However, if attackers continue to engage in micro-targeting of important or vulnerable groups, that will narrow the set of people who should be trained. This intervention is modeled on successful offline interactions, such as the use of partnerships with barber shops to educate black heterosexual men

[108] *Google and Microsoft Agree Steps to Block Abuse Images*, BBC NEWS (Nov. 18, 2013), http://www.bbc.com/news/uk-24980765 [https://perma.cc/V3T5-G4EG]; *Google Reveals 'Terrorism Video' Removals*, BBC NEWS (June 18, 2012), http://www.bbc.com/news/technology-18479137 [https://perma.cc/J9XA-FAFN]; *Remove Your Personal Information from Google*, GOOGLE, https://support.google.com/websearch/answer/2744324?hl=en [https://perma.cc/C2U4-NSLU] (last visited Oct. 14, 2019).

[109] *See* Loren Baker, *Google Explains Jew Watch Search Results*, SEARCH ENGINE J. (May 12, 2004), https://www.searchenginejournal.com/google-explains-jew-watch-search-results/552/ [https://perma.cc/H2YE-CTAG].

[110] *See Dropping the Bomb on Google*, WIRED (May 11, 2004, 12:00 PM), https://www.wired.com/2004/05/dropping-the-bomb-on-google/ [https://perma.cc/Y8J2-7R7J].

[111] *See* Kyle Cheney, *No, Clinton Didn't Start the Birther Thing. This Guy Did.*, POLITICO (Sept. 16, 2016, 3:55 PM), https://www.politico.com/story/2016/09/birther-movement-founder-trump-clinton-228304 [https://perma.cc/33SD-35UJ]; Marc Fisher et al., *Pizzagate: From Rumor, to Hashtag, to Gunfire in D.C.*, WASH. POST (Dec. 6, 2016), https://www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html [https://perma.cc/AFG6-SML3].

[112] *See* Lior J. Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919, 946–47 (2005).

about the risk of HIV/AIDS.[113] There are already organic examples, such as Reverend Brown's skepticism about the efforts of the purported Blacktivist group to generate conflict on the anniversary of the death of Freddie Gray.[114] Network theory suggests both a rationale for this intervention and a mechanism for it. It posits that certain people are key nodes that connect other people to one another, making them valuable for distributing accurate information or blocking disinformation.[115] Second, its analytical tools can enable platforms or researchers to assess the links among users, such as Facebook friendships, to discover who these key influencers are.[116]

Lastly, platforms could use whitelists—known sources of information that is generally accurate—to encourage users to share links to or data from reliable outlets. YouTube employed this technique to combat fake reviews about and attacks on actress Brie Larson during *Captain Marvel*'s run in theaters.[117] By designating the search term "Brie Larson" as newsworthy, YouTube altered its algorithm to prioritize results from trustworthy news sources.[118] This technique would need to be modified for platforms that are not centered around a search function. One possibility is that social networking sites could designate a set of key topics or terms, such as "Brie Larson" or "Pizzagate." If their users post links to articles from whitelisted sources, those posts would appear immediately. Posts from non-whitelisted sources might be delayed in appearing on the platform, either to enable further investigation of their accuracy or simply as a mechanism for encouraging users to rely on more trusted sources in their information distribution.

These tactics may be helpful in reducing the effects of disinformation, particularly on Internet platforms. However, it is important to acknowledge that the success of disinformation results from a larger set of social forces, particularly the breakdown of trust in previously respected gatekeepers.[119] These larger trends may or may not be reversible, but we cannot effectively combat disinformation without addressing them.

[113] *See* Tracey E. Wilson et al., *HIV Prevention for Black Heterosexual Men: The Barbershop Talk with Brothers Cluster Randomized Trial*, 109 AM. J. PUB. HEALTH 1131, 1131–35 (2019), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6611102/ [https://perma.cc/S3GW-M4JV].

[114] *See supra* Part I, at notes 32–34.

[115] *See* Strahilevitz, *supra* note 112, at 948–53.

[116] *See* BODINE-BARON ET AL., *supra* note 93, at 50–53.

[117] *See* Julia Alexander, *YouTube Fought Brie Larson Trolls by Changing Its Search Algorithm*, VERGE (Mar. 8, 2019, 12:12 PM), https://www.theverge.com/2019/3/8/18255265/brie-larson-youtube-captain-marvel-mcu-algorithm-review-bomb-trolls [https://perma.cc/7DKN-M48J].

[118] *Id.*

[119] *See* Verstraete & Bambauer, *supra* note 96, at 132.