SCIENCE AND DEMOCRACY

Protecting elections from social media manipulation

Rigorous causal analysis could help harden democracy against future attacks

By **Sinan Aral**[1,2,3] *and* **Dean Eckles**[1,2]

To what extent are democratic elections vulnerable to social media manipulation? The fractured state of research and evidence on this most important question facing democracy is reflected in the range of disagreement among experts. Facebook chief executive officer Mark Zuckerberg has repeatedly called on the U.S. government to regulate election manipulation through social media. But we cannot manage what we do not measure. Without an organized research agenda that informs policy, democracies will remain vulnerable to foreign and domestic attacks. Thankfully, social media's effects are, in our view, eminently measurable. Here, we advocate a research agenda for measuring social media manipulation of elections, highlight underutilized approaches to rigorous causal inference, and discuss political, legal, and ethical implications of undertaking such analysis. Consideration of this research agenda illuminates the need to overcome important trade-offs for public and corporate policy—for example, between election integrity and privacy. We have promising research tools, but they have not been applied to election manipulation, mainly because of a lack of access to data and lack of cooperation from the platforms (driven in part by public policy and political constraints).

Two recent studies commissioned by the U.S. Senate Intelligence Committee detail Russian misinformation campaigns targeting hundreds of millions of U.S. citizens during the 2016 presidential election. The reports highlight, but do not answer,

[1] Sloan School of Management, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [2] Institute for Data, Systems, and Society, MIT, Cambridge, MA, USA. [3] Manifest Investment Partners, Tiburon, CA, USA. Email: sinan@mit.edu

ILLUSTRATION: BROBEL DESIGN

Published by AAAS

whether social media manipulation may have influenced the outcome. Some experts argue that Russia-sponsored content on social media likely did not decide the election because Russian-linked spending and exposure to fake news (*1*, *2*) were small-scale. Others contend that a combination of Russian trolls and hacking likely tipped the election for Donald Trump (*3*). Similar disagreements exist about the UK referendum on leaving the European Union and recent elections in Brazil, Sweden, and India.

Such disagreement is understandable, given the distinctive challenges of studying social media manipulation of elections. For example, unlike the majority of linear television advertising, social media can be personally targeted; assessing its reach requires analysis of paid and unpaid media, ranking algorithms and advertising auctions; and causal analysis is necessary to understand how social media changes opinions and voting.

Luckily, much of the necessary methodology has already been developed. A growing body of literature illuminates how social media influences behavior. Analysis of misinformation on Twitter and Facebook (*4*, *5*), and randomized and natural experiments involving hundreds of millions of people on various platforms, have shown how social media changes how we shop, read, and exercise [e.g., (*6*, *7*)]. Similar methods can and should be applied to voting (*8*).

Research on election manipulation will be enabled and constrained by parallel policy initiatives that aim, for example, to protect privacy. Although privacy legislation may prohibit retention of consumer data, such data may also be critical to understanding how to harden our democracies against manipulation. To preserve democracy in the digital age, we must manage these tradeoffs and overcome multidisciplinary methodological challenges simultaneously.

MEASURING MANIPULATION

We propose a four-step research agenda for estimating the causal effects of social media manipulation on voter turnout and vote choice (see the figure). We also describe analysis of the indirect, systemic effects of social media manipulation on campaign messaging and the news cycle (see supplementary materials for further details).

Step 1: We must catalog exposures to manipulation, which we define as impressions (i.e., serving of an ad or message to a viewer) of paid and organic manipulative content (*9*) (e.g., false content intended to deceive voters, or even true content propagated by foreign actors, who are banned from participating in domestic political processes, with the intent of manipulating voters). To do so, we must evaluate the reach of manipulation campaigns and analyze the targeting strategies that distribute these impressions. For example, we need to know which text, image, and video messages were advertised, organically posted, and "boosted" through paid advertising, and on which platforms, as well as when and how each of these messages was shared and reshared by voters (*2*) and inauthentic accounts. Here, understanding social multiplier effects, or how individuals influence each other, will be essential, and the literature on peer effects in social networks describes how our peers change our behavior (*6*–*8*). The content of the messages should also be analyzed to assess the effectiveness of particular textual, image, and video content in changing opinions and behavior.

Much prior work on exposure to and diffusion of (mis)information has relied on proxies for exposure, such as who follows whom on social media (*2*, *4*), though some has also investigated logs of impressions, recognizing the role of algorithmic ranking and auctions in determining exposure [e.g., (*5*, *10*)]. Given prior work on the rapid decay of advertising effects, it is important to consider when these exposures occurred, as recent work suggests that exposure to misinformation may increase just prior to an election and wane immediately afterward (*2*).

Step 2: We must combine exposure data with data on voting behavior. Data about voter turnout in the United States are readily available in public records (e.g., registered voters' names, addresses, party affiliations, and when they voted). Prior work has matched social media accounts and public voting records using relatively coarse data (e.g., residences inferred from self-reported profile data and group-level, anonymous matching procedures) (*2*, *8*), in part because of privacy concerns, resulting in low match rates that limit statistical power and representativeness. This could be substantially improved, for example, by using the rich location data possessed by social media platforms, similar to that already sold and reused for marketing purposes (e.g., matching voter registrations with inferred home addresses based on mobile and other location data), rather than simply matching voters by name and age at the state level.

In contrast to turnout data, vote choices in the United States are secret and thus only measurable in aggregate (e.g., precinct-level vote totals and shares) or sparsely and indirectly through surveys (e.g., exit polls). Thus, exposure data would need to be aggregated, at the precinct, district, or state levels, before combining it with vote choice data, making it likely that estimates of voter turnout effects will be more precise than estimates of vote choice effects.

Experiments demonstrate that persuasive interventions can substantially affect voter turnout. But, when assessing turnout, it is important to remember that voting is habitual. Effective manipulation therefore likely requires targeting occasional voters in battleground regions. In social media, however, this type of targeting is possible and took place during the 2016 U.S. presidential election. Analysis of the precision of targeting efforts is essential to understanding voter turnout effects.

Influencing vote choice is more difficult because likely voters have strong prior beliefs. However, even the pessimistic litera-

Published by AAAS

ture on vote choice allows for substantial effects, especially when targeted messages change voters' beliefs. In a meta-analysis of 16 field experiments, Kalla and Broockman (*11*) report a wide 95% confidence interval (CI) of [−0.27%, 0.83%] for the effect of impersonal contact (e.g., mail, ads) on vote choice within 2 months of general elections, and larger, more significant effects in primaries and on issue-specific ballot measures. In Rogers and Nickerson (*12*), informing recipients in favor of abortion rights that a candidate was not consistently supportive of such rights had a 3.90% [95% CI: 1.16%, 6.64%] effect on reported vote choice. Such prior beliefs are predictable and addressable in manipulation campaigns through social media targeting and thus measurable in studies of the effectiveness of such manipulation.

A blueprint for empirical investigations of social media manipulation

ASSESS MESSAGE CONTENT AND REACH →	ASSESS TARGETING AND EXPOSURE →	ASSESS CAUSAL BEHAVIOR CHANGE →	ASSESS EFFECTS ON VOTING BEHAVIOR
How many messages spread?	Who was exposed to which messages?	How did messages change opinions and behavior?	How did opinion and behavior change alter voting outcomes?
Analysis of paid and organic information diffusion	Analysis of targeting and messaging exposure	Causal statistical analysis of opinion and behavior change	Counterfactual analysis of deviations from expected voting
Measure impressions through paid media and sharing	Evaluate targeting campaigns and impression distributions	Evaluate causal effects across individuals and segments	Measure deviations from expected voting behavior

Step 3: We must assess the effects of manipulative messages on opinions and behavior. This requires a rigorous approach to causal inference, as naïve, observational approaches would neglect the confounding factors that cause both exposure and voting behavior (e.g., voters targeted with such content are more likely to be sympathetic to it). Evaluations using randomized experiments have shown that observational estimates of social media influence without careful causal inference are frequently off by more than 100%. Effects of nonpaid exposures, estimated without causal inference, have been off by as much as 300 to 700%. Yet, causal claims about why social media messages spread are routinely made without any discussion of causal inference. Widely publicized claims about the effectiveness of targeting voters by inferred personality traits, as allegedly conducted by Cambridge Analytica, were not based on randomized experiments or any other rigorous causal inference and therefore plausibly suffer from similar biases.

To credibly estimate the effects of misinformation on changes in opinions and behaviors, we must change our approach and embrace causal inference. We must analyze similar people exposed to varying levels of misinformation, perhaps due to random chance or explicit randomization by firms and campaigns. Fortunately, there are many, until-now largely ignored, sources of such random variation. For example, Facebook and Twitter constantly test new variations on their feed ranking algorithms, which cause people to be exposed to varying levels of different types of content. Some preliminary analysis suggests that an A/B test run by Facebook during the 2012 U.S. presidential election caused over 1 million Americans to be exposed to more "hard news" from established sources, affecting political knowledge, policy preferences, and voter turnout (*10*). Most of these routine experiments are not intended specifically to modulate exposure to political content, but recent work has illustrated how the random variation produced by hundreds or thousands of routine tests, of the kind these platforms conduct every day, can be used to estimate the effects of exposure to such content (*13*). Such experiments could facilitate measurement of both direct effects (e.g., effects of manipulative content on recipients) and indirect "spillover" effects (e.g., word of mouth from recipients to peers), though other methods for estimating the latter also exist (*6*–*8*).

One important challenge is that statistical precision is often inadequate to answer many questions about effects on voter behavior. For example, randomized experiments conducted by Facebook in the 2010 and 2012 U.S. elections only barely detected effects on turnout—even though the estimated effects imply that a minimal intervention caused hundreds of thousands of additional votes to be cast [e.g., (*8*)]. The lack of statistical precision in those studies arose in part because only about a tenth of users were uniquely matched to voter records, which, as we note, could be improved upon. Furthermore, unlike television advertising, much less of the as-good-as-random variation in exposure to social media may be within, not between, geographic areas, making effects on aggregate vote shares more difficult to detect. Such imprecision can be misleading, suggesting that online advertising does not work simply because the effects were too small to detect in a given study (*14*), even though the results were consistent with markedly low costs per incremental vote, making engagement in such campaigns economically rational.

Step 4: We must compute the aggregate consequences of changes in voting behavior for election outcomes. To do so, we would combine summaries of individual-level counterfactuals (i.e., predicted voter behavior with and without exposure) with data on the abundance of exposed voters by geographic, demographic, and other characteristics in specific elections. This would enable estimates and confidence intervals for vote totals in specific states or regions if a social media manipulation campaign had not been conducted. Although some of these confidence intervals will include vote totals that do or do not alter the winner in a particular contest, the ranges of counterfactual outcomes would still be informative about how such manipulation can alter elections. Although it remains to be seen exactly how precise the resulting estimates of the effects of exposure to misinformation would be, even sufficiently precise and carefully communicated null results could exclude scenarios currently posited by many commentators.

Research should also address the systemic effects of social media manipulation, like countermessaging and feedback on the news cycle itself. Countermessaging could be studied in, for example, the replies to and debunking of fake news on Facebook and Twitter (*4*, *5*) and whether the emergence of fake stories alters the narrative trajectories of messaging by campaigns or other interested groups. Feedback into the news cycle could be studied by examining the causal impact of manipulation on the topical content of news coverage. For example, Ananya Sen and Pinar Yildirim have used as-good-as-random variation in the weather to show that more viewership to particular news stories causes publishers to write more stories on those topics. A similar approach could determine whether attention to misinformation alters the topical trajectory of the news cycle.

We believe near-real-time and ex post analysis are both possible and helpful. The bulk of what we are proposing is ex post analysis of what happened, which can then be used to design platforms and policy to prevent future manipulation. The pace at which voting data (whether in primaries or general elections) become available is a key limitation. But real-time detection of manipulation efforts and reaction to them could also be designed, similar to tactics in digital advertising that estimate targeting models offline and then implement real-time bidding based on those estimates. Experimental analysis of the effect of social media on behavior change can be spun up and conducted by the platforms in a matter of days and analyzed in a week.

LEGAL, ETHICAL, AND POLITICAL IMPLICATIONS

We have described what a rigorous analysis of social media manipulation would entail, but have also assumed that the data required to conduct it are available for analysis. But does the social media data that we describe above, especially data about the content that individuals were exposed to, exist retrospectively or going forward? Social media companies routinely log what users are exposed to for research and retraining algorithms. But current regulatory regimes disincentivize the lossless retention of this data. For example, the European Union's General Data Protection Regulation (GDPR) encourages firms to comply with user requests to delete data about them, including content that they have posted. An audit by the office of the Irish Data Protection Commissioner caused Facebook to implement similar policies in 2012. Thus, without targeted retention, it may be difficult for firms to accurately quantify exposures for users who deleted their accounts or were exposed to content deleted by others. We should recognize that well-intentioned privacy regulations, though important, may also impede assessments like the one that we propose. Similarly, proposed legislation in the United States (the DETOUR Act) could make many routine randomized experiments by these firms illegal (Senate Bill 1084), making future retrospective analyses more difficult and, of course, making ongoing efforts by those firms to limit such manipulation less data-driven.

Even if such data are available, it is not obvious that we should accept world governments demanding access to or analyses of those data to quantify the effects of speech in elections. Although we suggest that linking datasets could be achieved using rich location data routinely used for marketing, such use may be reasonably regarded as data misuse. Thus, we do not unconditionally advocate the use of any and all existing data for the proposed analyses. Instead, privacy-preserving methods for record linkage and content analysis, such as differential privacy (*15*), could help manage trade-offs between the need for privacy and the need to protect democracy.

Hardening democracies to manipulation will take extraordinary political and commercial will. Politicians in the United States, for example, may have countervailing incentives to support or oppose a postmortem on Russian interference, and companies like Facebook, Twitter, and Google face pressure to secure personal data. Perhaps this is why Social Science One, the forward-looking industry–academic partnership working to provide access to funding and Facebook data to study the effects of social media on democracy, faced long delays in securing access to any data, and why its most recent release does not include any data relevant to a postmortem on Russian interference in the 2016 or 2018 elections in the United States. Moreover, this cannot just be about any single company or platform. Comprehensive analysis must include Facebook, Twitter, YouTube, and others. Perhaps only mounting pressure from legislators and the public will empower experts with the access they need to do the work that is required.

"...begin a public discussion of the trade-offs between privacy, free speech, and democracy..."

Research collaborations with social media platforms, like that being undertaken by Social Science One, can facilitate access to important data for understanding democracy's vulnerability to social media manipulation. We hope the realization that the analysis we propose is bigger than any one election and essential to protecting democracies worldwide will help overcome partisanship and myopic commercial interests in making the necessary data available, in privacy-preserving ways.

However, it is important to note that prior work has linked social media messaging to validated voting, both with the assistance of the social media platforms (*8*) and without it (*2*). Although collaboration with the platforms is preferable, it is not the only way to assess manipulation. In the absence of commercial or governmental support for postmortems on past elections, active analysis of ongoing information operations, conducted according to the framework that we propose, is a viable and valuable alternative. A detailed understanding of country-specific regulations and election procedures is necessary for robust analysis of the effects of social media manipulation on democracies worldwide.

Our suggested approach emphasizes precise causal inference, but this should be complemented with surveys, ethnographies, and analysis of observational data to understand the mechanisms through which manipulation can affect opinions and behavior.

Achieving a scientific understanding of the effects of social media manipulation on elections is an important civic duty. Without it, democracies remain vulnerable. The sooner we begin a public discussion of the trade-offs between privacy, free speech, and democracy that arise from the pursuit of this science, the sooner we can realize a path forward. ■

REFERENCES AND NOTES

1. H. Allcott, M. Gentzkow, *J. Econ. Perspect.* **31**, 211 (2017).
2. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, *Science* **363**, 374 (2019).
3. K. H. Jamieson, *Cyberwar: How Russian Hackers and Trolls Helped Elect a President* (Oxford Univ. Press, 2018).
4. S. Vosoughi, D. Roy, S. Aral, *Science* **359**, 1146 (2018).
5. A. Friggeri, L. A. Adamic, D. Eckles, J. Cheng, in *Proceedings of the International Conference on Web and Social Media* (Association for the Advancement of Artificial Intelligence, 2014).
6. S. Aral, D. Walker, *Science* **337**, 337 (2012).
7. S. Aral, C. Nicolaides, *Nat. Commun.* **8**, 14753 (2017).
8. R. M. Bond *et al.*, *Nature* **489**, 295 (2012).
9. A. Guess, B. Nyhan, J. Reifler, Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign (European Research Council, 2018).
10. S. Messing, *Friends that Matter: How Social Transmission of Elite Discourse Shapes Political Knowledge, Attitudes, and Behavior*, Ph.D. thesis, Stanford University (2013).
11. J. L. Kalla, D. E. Broockman, *Am. Polit. Sci. Rev.* **112**, 148 (2018).
12. T. Rogers, D. Nickerson, Can inaccurate beliefs about incumbents be changed? And can reframing change votes? HKS Working Paper no. RWP13-018 (2013).
13. A. Peysakhovich, D. Eckles, Learning causal effects from many randomized experiments using regularized instrumental variables, in *Proceedings of the 2018 World Wide Web Conference* (International World Wide Web Conferences Steering Committee, 2018), pp. 699–707.
14. D. E. Broockman, D. P. Green, *Polit. Behav.* **36**, 263 (2014).
15. C. Dwork, Differential privacy: A survey of results, in *Proceedings of the International Conference on Theory and Applications of Models of Computation* (Springer, 2008).

ACKNOWLEDGMENTS

We thank A. J. Berinsky and B. Nyhan for comments. S.A. has financial interest in Alibaba, Google, Amazon, and Twitter. S.A. was a Scholar in Residence at the *New York Times* in 2013 and visiting researcher at Microsoft in 2016. S.A. has received research funding from *The Boston Globe* and speaking fees from Microsoft. S.A. is an inventor on a related patent pending. D.E. has financial interest in Facebook, Amazon, Google, and Twitter. D.E. was a consultant at Microsoft in 2018. D.E. was an employee and consultant at Facebook from 2010 to 2017. D.E. has recently received funding from Amazon. D.E.'s attendance at conferences has recently been funded by DARPA, Microsoft, and Technology Crossover Ventures. D.E. is an inventor on a related patent, which is assigned to Facebook. S.A. and D.E. contributed equally to this work.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6456/858/suppl/DC1

10.1126/science.aaw8243

Science

Protecting elections from social media manipulation

Sinan Aral and Dean Eckles

Science **365** (6456), 858-861.
DOI: 10.1126/science.aaw8243

science.sciencemag.org/content/365/6456/858/suppl/DC1

Supplementary Materials for

Protecting elections from social media manipulation

Sinan Aral* and Dean Eckles

*Corresponding author. Email: sinan@mit.edu

Published 30 August 2019, *Science* **365**, 858 (2019)
DOI: 10.1126/science.aaw8243

This PDF file includes:

Supplementary Text
References

Supplementary Text

Here we provide additional references supporting key arguments in the main text.

1. DISAGREEMENT AMONGST EXPERTS

Varied statements on the likelihood that Russian-sponsored social media content substantially affected voting behavior appear in many places, including Allcott & Gentzkow (2017), Guess, Nagler, and Tucker (2019), Guess, Nyhan, and Reifler (2018), Jamieson (2018), and Sides, Tesler, and Vavreck (2018).

Disagreement among experts about whether social media manipulation has or could affect the results of elections stems from differing beliefs about (a) the likely reach and scope of misinformation campaigns and (b) the likely effects of social media manipulation on voter turnout and vote choice.

2. THE REACH AND SCOPE OF MISINFORMATION CAMPAIGNS

While some research estimates that Russian misinformation, for example, reached hundreds of millions of people on social media during the 2016 U.S. Presidential election (DiResta et al., 2018; Howard et al., 2018), others contend the reach and scope of exposures was small, concentrated and selective (Allcott & Gentzkow, 2017; Grinberg et al., 2019, Guess, Nagler & Tucker, 2019; Guess, Nyhan & Reifler, 2018).

3. EFFECTS ON VOTER TURNOUT AND VOTE CHOICE

There is also disagreement on the effectiveness of social media persuasion and whether it could be substantial enough to tip an election. Here it is important to distinguish the likely effects of manipulation on voter turnout and vote choice.

With regard to vote choice, some meta-analytic reviews suggest the effects of impersonal contact (e.g., mailing, TV and digital advertising) on vote choice in elections are very small. For example, Kalla and Broockman (2017) conclude that “the best estimate of the size of persuasive effects [i.e., effects of advertising on vote choice] in general elections in light of our evidence is zero.” However, there remains substantial uncertainty and heterogeneity in their estimates, reflected in the confidence interval reported in the main text, from their Figure 4b, for the meta-analytic effect of impersonal contact within two months of election day. Kalla and Broockman (2017) also find significant meta-analytic effects on vote choice in primaries, issue specific ballot measures and when campaigns target persuadable voters, suggesting the possibility that manipulation is effective in changing vote choices when they are issue specific and targeted. We also note that the social media manipulation we have observed to date has typically been issue specific and targeted, similarly to the randomized intervention cited in the main text (Rogers & Nickerson, 2013).

Furthermore, social media manipulation does not have to affect vote choice to tip an election. Effects on voter turnout, if well targeted, could be substantial enough to change an overall result. The meta-analytic assessments of voter turnout point to much more substantial effects. For example, the meta-analysis by Green et al. (2013) estimates that direct mailings with social pressure generate an average increase in voter turnout of 2.9% (95% CI = 2.7%-3.0%), canvassing generates an average increase of 2.5% (95% CI = 1.8%-3.3%) and volunteer phone banks generate an average increase of 2% (95% CI = 1.3%-2.6%). Dale and Strauss (2009) estimate the voter turnout effect of text messages to be 4.1% and there is also evidence that personalized emails create substantial voter turnout effects (Davenport 2012; Malhotra et al., 2012). The only studies of voter turnout effects from social media messaging estimate that hundreds of thousands of additional votes were cast as a result of social media messages (Bond et al., 2012; Jones et al., 2017).

4. MEASURING EXPOSURE

Much prior work on exposure to and diffusion of (mis)information has relied on proxies for exposure. However, some work by researchers at Facebook (Bakshy et al., 2012a,b; Bakshy, Eckles & Bakshy, 2017; Friggeri et al., 2014; Messing & Adamic, 2015; Messing, 2013) has made use of detailed data about impressions (delivery of content to the users' device), including information about what content was actually displayed to a user for at least a minimum period of time, thus making use of measures now in widespread use in digital advertising.

5. LINKING EXPOSURE TO VOTING DATA

Voter turnout in the United States is a matter of public record, so voter records including data about individuals' turnout is widely used by campaigns and researchers. Bond et al. (2012) linked Facebook accounts to voter turnout data to estimate effects of a randomized intervention. They did this matching using limited information, apparently because of privacy concerns, as articulated in a companion paper about privacy-preserving record linkage (Jones et al., 2013). A subsequent experiment used similarly coarse data for record linkage and resulted in similarly low unique match rates (Jones et al., 2017).

6. BIAS IN NAÏVE OBSERVATIONAL STUDIES

Aral, Muchnik, and Sundararajan (2009) compare the results of naïve observational methods to counterfactual methods based on matching and find naïve methods overestimate the effects of non-paid exposure to behavior of friends in an online social network by 300-700%. At least since LaLonde (1986), researchers have used randomized experiments as a "gold standard" with which to evaluate other, observational methods, like matching. In the context of the diffusion of (mis)information, Eckles and Bakshy (2017) validate the methods used in Aral, Muchnik and Sundararajan (2009) by comparing the results of a large field experiment on Facebook to analyses of matching methods. They find naïve methods overestimate the effects of non-paid exposure to content shared by friends by over 300% and demonstrate that matching can reduce this bias by up to 80–100%.

Non-experimental methods for estimating effects of paid exposure (digital advertising) have also performed poorly when similarly evaluated. Gordon et al. (2018) used randomized experiments to show observational estimates of social media influence, without careful causal inference, are frequently off by over 100%.

Similar confounding is plausibly present in widely-publicized claims (Matz et al., 2018) about the effectiveness of targeting ads according to inferred personality traits (Eckles, Gordon & Johnson, 2018).

7. QUASI-EXPERIMENTAL METHODS FOR ESTIMATING EFFECTS ON VOTING BEHAVIORS

Several studies have exploited a mismatch between borders of competitive electoral districts and borders of regions for marketing purposes to study effects of advertising, including Huber and Arceneaux (2007) and more recent work (Spenkuch & Toniatti, 2018; Wang, Lewis & Schweidel, 2018). However, targeting of digital advertising is less restricted to such borders compared with traditional, linear television advertising, making this source of plausibly exogenous variation in exposure largely inapplicable in the digital arena.

8. ROUTINE EXPERIMENTATION BY PLATFORMS

Internet companies are engaged in continual experimentation, with the most prominent firms starting hundreds of experiments each week (Bakshy, Eckles & Bernstein, 2014; McAfee, A., & Brynjolfsson, 2012; Varian, 2016). Even a single part of a product, such as the algorithm for ranking search results or a feed of content shared by others (e.g., News Feed), might be modulated in hundreds or thousands of experiments over the course of a campaign (Kohavi & Thomke, 2017; Peysakhovich & Eckles, 2018). Most of these experiments are not designed for studying exposure to political content, with the exception of, e.g., Messing (2013), covered in Sifry (2014). However, such experiments can be key inputs for recently developed methods for high-dimensional instrumental variables regression (e.g., Kang et al., 2016; Belloni et al., 2017; Peysakhovich & Eckles, 2018; Guo et al., 2018).

9. INDIRECT EFFECTS

Like other content on social media, effects of Russian-sponsored content may occur indirectly via diffusion of the content and further social contagion (cf. Nickerson, 2008; Bond et al., 2012; Jones et al., 2017). There has been substantial recent development of methods for estimation of such “spillover” effects in networks (e.g. Aronow, 2012; Aronow & Samii, 2017; Athey, Eckles & Imbens, 2018), with empirical work in online social networks making use of both designed experiments (e.g., Aral & Walker, 2011, 2012, 2014; Bakshy et al., 2012a,b; Eckles, Kizilcec & Bakshy, 2016; Huang et al., 2019; Muchnik, Aral & Taylor, 2013), natural quasi-experiments (e.g., Aral & Nicolaides, 2017; Aral & Zhao, 2018) and other causal inference methods (e.g., Aral, Muchnik & Sundararajan, 2009; Eckles & Bakshy, 2017).

There is also reason to believe that there are temporal spillovers, with the effects of persuasive messaging in one election spilling over into future elections (Gerber, Green & Shachar, 2003; Davenport et al., 2010; Bedolla & Michelson, 2012). This is consistent with the idea that voting is habitual (e.g., Plutzer, 2002; Gerber, Green & Shachar, 2003) and that messaging can affect voting habits.

To the extent that social media activity is subsequently covered by the news media, this might result in effects on the voting behavior of those who are not directly exposed on social media. This would require different empirical strategies for credible causal inference, such as in Sen and Yildirim (2016).

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211-36.

Aral, S. & Nicolaides, C. (2017). Exercise contagion in a global social network. Nature Communications, 8(14753): 1-8.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of the National Academy of Sciences, 106(51), 21544-21549.

Aral, S., & Walker, D. (2014). Tie strength, embeddedness & social influence: A large-scale networked experiment. Management Science, 60(6): 1352 - 1370.

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. Science, July 20: 337-341.

Aral, S. & Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. Management Science, 57(9); September: 1623-1639.

Aral, S., & Zhao, M. (2018). Social media and online news consumption. MIT Working Paper.

Aronow, P. M., & Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. The Annals of Applied Statistics, 11(4), 1912-1947.

Athey, S., Eckles, D., & Imbens, G. W. (2018). Exact p-values for network interference. Journal of the American Statistical Association, 113(521), 230-240.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), 9216-9221.

Barrientos, A. F., Reiter, J. P., Machanavajjhala, A., & Chen, Y. (2018). Differentially private significance tests for regression coefficients. Journal of Computational and Graphical Statistics, forthcoming.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012a). The role of social networks in information diffusion. In Proceedings of the 21st international conference on World Wide Web (pp. 519-528). ACM.

Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012b). Social influence in social advertising: evidence from field experiments. In Proceedings of the 13th ACM conference on Electronic Commerce. ACM.

Bakshy, E., Eckles, D., & Bernstein, M. S. (2014). Designing and deploying online field experiments. In Proceedings of the 23rd international conference on World wide web (pp. 283-292). ACM.

Bakshy, B., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science 348, 1130–1132 (2015)

Bedolla, L. G., & Michelson, M. R. (2012). Mobilizing inclusion: Transforming the electorate through get-out-the-vote campaigns. Yale University Press.

Belloni, A., Chernozhukov, V., Fernández-Val, I., & Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. Econometrica, 85(1), 233-298.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120). ACM.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences, 114(28), 7313-7318.

Broockman, D. E., & Green, D. P. (2014). Do online advertisements increase political candidates' name recognition or favorability? Evidence from randomized field experiments. Political Behavior, 36(2), 263-289.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, 90(10), 60-68.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. Science, 341(6146), 647-651.

Dale, A. & Strauss, A. (2009). Don't forget to vote: text message reminders as a mobilization tool. American Journal of Political Science, 53(4), pp. 787-804.

Davenport, T.C. (2012) Unsubscribe: The effects of peer-to-peer email on voter turnout – results from a field experiment in the June 6, 2006, California primary election. Unpublished manuscript (Yale University)

Davenport, T. C., Gerber, A. S., Green, D. P., Larimer, C. W., Mann, C. B., & Panagopoulos, C. (2010). The enduring effects of social pressure: Tracking campaign experiments over a series of elections. Political Behavior, *32*(3), 423-430.

DeVries, J. V., Singer, N., Keller, M. H., & Krolik, A. (2018, December 10). Your apps know where you were last night, and they're not keeping it secret. New York Times.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J., & Johnson, B. (2018). The Tactics & Tropes of the Internet Research Agency. Report, New Knowledge. https://www.newknowledge.com/articles/the-disinformation-report/

Dwork, C. (2008). Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models of Computation. Springer.

Eckles, D., & Bakshy, E. (2017). Bias and high-dimensional adjustment in observational studies of peer effects. Working paper. https://arxiv.org/abs/1706.04692.

Eckles, D., Gordon, B. R., & Johnson, G. A. (2018). Field studies of psychologically targeted ads face threats to internal validity. Proceedings of the National Academy of Sciences, 201805363.

Eckles, D., Kizilcec, R. F., & Bakshy, E. (2016). Estimating peer effects in networks with peer encouragement designs. Proceedings of the National Academy of Sciences, 113(27), 7316-7322.

Friggeri, A., Adamic, L. A., Eckles, D., & Cheng, J. (2014). Rumor cascades. In Proceedings of the International Conference on Web and Social Media. AAAI.

Gerber, A. S., Green, D. P., & Shachar, R. (2003). Voting may be habit-forming: evidence from a randomized field experiment. American Journal of Political Science, 47(3), 540-550.

Gerber, A. S., Gimpel, J. G., Green, D. P., & Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. American Political Science Review, 105(1), 135-150.

Gerber, A.S., Green, D.P. & Shachar, R. (2003). Voting may be habit-forming: evidence from a randomized field experiment. American Journal of Political Science, 47, pp. 540-550.

Gordon, B. R., Zettelmeyer, F., Bhargava, N., & Chapsky, D. (2018). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3162023.

Green, D. P., McGrath, M. C., & Aronow, P. M. (2013). Field experiments and the study of voter turnout. Journal of Elections, Public Opinion and Parties, 23(1), 27-48.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. Science 363(6425), 374-378.

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, 5(1), eaau4586.

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. European Research Council.

Guo, Z., Kang, H., Tony Cai, T., & Small, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(4), 793-815.

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). The IRA, Social Media and Political Polarization in the United States, 2012-2018. Report, University of Oxford. https://www.graphika.com/ssci-report/

Howard, P. N., Kollanyi, B., Bradshaw, S., & Neudert, L. M. (2018). Social media, news and political information during the US election: Was polarizing content concentrated in swing states?. Working paper. https://arxiv.org/abs/1802.03573

Huang, S., Aral, S., Brynjolfsson, E., & Hu, J. (2019). Social advertising effectiveness across products: A large-scale field experiment. Working paper, MIT.

Huber, G. A., & Arceneaux, K. (2007). Identifying the persuasive effects of presidential advertising. American Journal of Political Science, 51(4), 957-977.

Jamieson, K. H. (2018). Cyberwar: How Russian Hackers and Trolls Helped Elect a President. Oxford University Press.

Jones, J. J., Bond, R. M., Fariss, C. J., Settle, J. E., Kramer, A. D., Marlow, C., & Fowler, J. H. (2013). Yahtzee: An anonymized group level matching procedure. PloS one, 8(2), e55760.

Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., & Fowler, J. H. (2017). Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 US presidential election. PloS one, 12(4), e0173851.

Kalla, J. L., & Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. American Political Science Review, 112(1), 148-166.

Kang, H., Zhang, A., Cai, T. T., & Small, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. Journal of the American Statistical Association, 111(513), 132-144.

Kohavi, R., & Thomke, S. H. (2017). The surprising power of online experiments: Getting the most out of A/B and other controlled tests. Harvard Business Review 95(5), 74–82.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. The American Economic Review, 76(4), 604-620.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The science of fake news. Science, 359(6380), 1094-1096.

Malhotra, N., Michelson, M.R., & Valenzuela, A.A. (2012). Emails from official sources can increase turnout. Quarterly Journal of Political Science, 7, pp. 321-332.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences, 114(48), 12714-12719.

Messing, S. (2013). Friends that Matter: How Social Transmission of Elite Discourse Shapes Political Knowledge, Attitudes, and Behavior (Doctoral dissertation, Stanford University).

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. Communication Research, 41(8), 1

Nickerson, D. W. (2008). Is voting contagious? Evidence from two field experiments. American Political Science Review, 102(1), 49-57.

Office of Irish Data Protection Commissioner (OIDPC). (2011). Facebook Ireland Ltd.: Report of Audit.

Peysakhovich, A., & Eckles, D. (2018). Learning causal effects from many randomized experiments using regularized instrumental variables. In Proceedings of the 2018 World Wide Web Conference on World Wide Web (pp. 699-707). International World Wide Web Conferences Steering Committee.

Provost, F., Dalessandro, B., Hook, R., Zhang, X., & Murray, A. (2009). Audience selection for on-line brand advertising: privacy-friendly social network targeting. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 707-716). ACM.

Plutzer, E., (2002). Becoming a habitual voter: Inertia, resources, and growth in young adulthood. American Political Science Review, 96(1), 41-56.

Rogers, R., & Nickerson, D. (2013). Can inaccurate beliefs about incumbents be changed? And can reframing change votes?. HKS Working Paper No. RWP13-018.

Sen, A., & Yildirim, P. (2016). Clicks bias in editorial decisions: How does popularity shape online news coverage? Working paper. https://ssrn.com/abstract=2619440 or http://dx.doi.org/10.2139/ssrn.2619440

Senate Bill 1084 (2019). Deceptive Experiences To Online Users Reduction Act, 116th Congress.

Shapiro, B. T. (2018). Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. Journal of Political Economy, 126(1), 381-437.

Sides, J., Tesler, M., & Vavreck, L. (2018). Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America. Princeton University Press.

Sifry, M. (2014, October 31). Facebook wants you to vote on Tuesday. Here's how it messed with your feed in 2012. Mother Jones. https://www.motherjones.com/politics/2014/10/can-voting-facebook-button-improve-voter-turnout/

Spenkuch, J. L., & Toniatti, D. (2018). Political advertising and election results. The Quarterly Journal of Economics, 133(4), 1981-2036.

Varian, H. (2016). Intelligent technology. Finance and Development, 53, 3 (2016).

Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. Science, 359(6380): 1146-1151.

Wang, Y., Lewis, M., & Schweidel, D. A. (2018). A border strategy analysis of ad source and message tone in senatorial campaigns. Marketing Science, 37(3), 333-355.