Supplementary Materials for

Protecting Elections from Social Media Manipulation

Sinan Aral,*† Dean Eckles*

*These authors contributed equally to this work.
†Corresponding author. Email: sinan@mit.edu

Published 5 September 2019, Science XXX, XXX (2019)
DOI: XX.XXX/science.aaoXXXX

This PDF Includes:

Supporting Materials
References

Supporting Materials

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.