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


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


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


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


