

Going Micro to Go Negative?

Targeting Toxicity using Facebook and Instagram Ads

Fabio Votta

University of Amsterdam

f.a.votta@uva.nl

Arman Noroozian

University of Amsterdam

Tom Dobber

University of Amsterdam

Natali Helberger

University of Amsterdam

Claes de Vreese

University of Amsterdam

Abstract

Spreading uncivil negative campaign messages is a “high-risk, high reward” campaign strategy since certain voters are more likely to be swayed by negative messaging whereas other voters are more inclined to feel sympathy with the attacked. Due to its risks, campaigns may attempt to outsource their uncivil ads to outside groups thus distancing themselves from the negativity and potentially avoiding any backlash. But at a time when advertising platforms boast of their ability to deliver ads to highly targeted audiences, uncivil negative ads could also be optimized to narrowly target citizens to which they are more likely to appeal. To study whether such optimizations are occurring, we retrieve all online advertisements that were placed on Facebook platforms (incl. Instagram) in the seven months prior to the US 2020 election. We perform multilevel ordinal regressions and find that ads from official political campaigns are more likely to be toxic when targeted

at a narrower audience, whereas “dark money” outside groups (like super PACs and non-profits) are more likely to target broad audiences with their toxicity. In addition, we find that ads from outside groups are more likely to be toxic. We discuss the findings in light of this evidence and reflect upon future research regarding microtargeting negative messages on online platforms such as Facebook and Instagram.

Keywords: political communication, microtargeting, negative campaigning, toxicity, ad library

Spreading uncivil negative campaign messages is a “high-risk, high reward” campaign strategy since certain voters are more likely to be swayed by negative messaging whereas other voters are more inclined to feel sympathy with the attacked. Due to its risks, campaigns may attempt to outsource their uncivil ads to outside groups thus distancing themselves from the negativity and potentially avoiding any backlash. But at a time when advertising platforms boast of their ability to deliver ads to highly targeted audiences, uncivil negative ads could also be optimized to narrowly target citizens to which they are more likely to appeal. To study whether such optimizations are occurring, we retrieve all online advertisements that were placed on Facebook platforms (incl. Instagram) in the seven months prior to the US 2020 election. We perform multilevel ordinal regressions and find that ads from official political campaigns are more likely to be toxic when targeted at a narrower audience, whereas “dark money” outside groups (like super PACs and non-profits) are more likely to target broad audiences with their toxicity. In addition, we find that ads from outside groups are more likely to be toxic. We discuss the findings in light of this evidence and reflect upon future research regarding microtargeting negative messages on online platforms such as Facebook and Instagram.

Introduction

The 2020 US presidential election has seen the importance of data-driven digital campaigns skyrocket. Fueled by a pandemic that forced campaign activities away from personal interactions and towards the digital sphere, spending on Facebook advertisements by the Trump and Biden campaign in 2020 each individually outclassed what the Trump and Clinton campaign had spend on Facebook ads in 2016 *combined* (Baum, 2020; Wagner, 2017).

This extreme increase in resources spent on data-driven digital campaigns raises important questions about the practice of *political microtargeting*, which allows tailored political messages to be delivered to specific audiences.

In particular, two events have sparked increasing concern regarding political microtargeting: 1.) the *2018 Cambridge Analytica scandal* in which the company produced ads that reportedly sought to manipulate people's voting intentions by targeting psychological profiles (Confessore, 2018); and 2.), the activities of the Russia-based Internet Research Agency (IRA) that placed targeted Facebook ads that were meant to elicit anger and fear to influence the US 2016 election (Vargo & Hopp, 2020).

Sobieraj and Berry, 2011 argue that political incivility “thrives in a narrowcasting environment” because actors in such an environment are able to “reach out to smaller and more homogeneous audiences and can afford to offend” because they don't have to fear backlash from individuals that find incivility objectionable. In their work, Sobieraj and Berry, 2011 were talking about the fragmented hyper-partisan U.S. media landscape in particular but the implications are even stronger for political microtargeting: using the detailed ad targeting capabilities of social media platforms like Facebook it is possible for advertisers to deliver negative messages to the “right voter” by reducing the risk of exposing individuals to ads that they find objectionable. In this way, political microtargeting may be used as a strategic choice to maximize the benefits of ads and reduce the risks of mismatching messages with receivers in ways that could backfire on the sponsor of such messages (Nai & Maier, 2020).

This study uses Facebook ad library data to learn about the strategic use of targeting smaller subsections of society. In particular, we investigate whether and how political microtargeting techniques are used to deliver campaign ads with toxic language in order to mitigate the risks that come with such a strategy. More specifically, we answer the following question:

To what extent is political microtargeting used by political advertisers to deliver toxic campaign messages?

Given that we are dealing with millions of online political advertisements, a computational approach of retrieving, wrangling, and analyzing the data is crucial to extract insights and answer our research question (Theocharis & Jungherr, 2021; van Atteveldt & Peng, 2018). We employ the Perspective API to measure toxicity in ads and validate the output with human coders (Perspective API, 2021; Wakabayashi, 2017). We further employ Mozilla DeepSpeech to transcribe the audio in ads (Hannun et al., 2014), and make

use of Google Cloud Vision API to extract texts from images using Optical Character Recognition (OCR) (Google Cloud Vision API, 2022).

Theoretical background

Going Micro

Social media platforms like Facebook afford advertisers a large audience as well as sophisticated online advertising tools. Political campaigns benefit from these affordances, both in terms of the sheer audience size that they can reach, as well as the wide range of targeting options available to them to reach out to specific groups (Dommett, 2019). These options include highly granular criteria such as locations, socio-demographic characteristics like age, gender, education levels and marital status, or even information about political preferences, as well as many other personal data and combinations thereof.

Microtargeting involves two distinct steps: first, specific target groups are defined using some input data about the desired audience. This input can be defined on the basis of targeting criteria that the platform provides, or be based on data owned by the advertisers themselves: curated lists of e-mails, telephone numbers, data scraped from public records, or purchased from commercial data brokers (Kreiss, 2016). Social media companies like Facebook then internally match this data with their user base to help create *custom* or *lookalike audiences* that share similar characteristics as the intended target audience (Ghosh et al., 2019). Next, targeted individuals receive customized messages that suit their personal circumstances to accomplish a political objective such as securing votes or donations from the targeted subgroup.

One of the main reasons why political parties may want to adopt microtargeting techniques when campaigning is that parties may see it as more *efficient* (Barbu, 2014; Zuiderveen Borgesius et al., 2018). With limited resources it makes sense that money is spend where it 'matters most', and microtargeting promises to political advertisers exactly that.

Broader audiences are likely to include people who are not susceptible to your message and spending money to show messages to these groups of people could be seen as a waste of campaign resources (Metcalf et al., 2018). In contrast, microtargeting allows advertisers to focus their attention and time on critical target audiences who can be sought out directly and more frequently if wanted. Further, advertisers may make use of microtargeting

because it might be more *effective*: evidence suggests that personality or susceptibility tailored political ads are indeed effective in swaying the opinions of exposed citizens (Dobber et al., 2020; Krotzek, 2019; Lavigne, 2020; Zarouali et al., 2020).

In the context of this paper, we will use a broad definition of micro-targeting, one which focuses on its ability to target a specific limited group of people by ‘narrowcasting’ a message, in opposition to ‘broadcasting’ a message to a bigger audience size (Raynauld & Turcotte, 2018; Sobieraj & Berry, 2011).

Going Negative

A commonplace political campaign strategy in countries around the globe is the practice of ‘going negative’ (Valli & Nai, 2020), a strategy that seeks to place a political competitor or rival in a negative light in order to enhance one’s image and sway voters to one’s own side (Geer, 2006). A key benefit for campaigns to engage in this practice is the human tendency to value information accompanied with negative tone over neutral- or positively presented information, or a so-called *negativity bias* (Fiske, 1980; Hilbig, 2009; Rozin & Royzman, 2001). Due to this information-processing asymmetry, campaign messages that attack opponents are expected to be more effective than positive messages in swaying voters (Meffert et al., 2006). However, scholars have yet to come to an agreement about the effectiveness of negative campaigning and the evidence for it in meta-analyses remains somewhat thin (Lau & Rovner, 2009; Lau et al., 2007).

An important consideration for assessing the supposed effectiveness of negative campaigning is that it has also been found to make people feel more sympathetic towards the target of the attack and to worsen attitudes towards the attacker instead (Walter & van der Eijk, 2019). Researchers first documented this *backlash effect* in the 1980ies (Garramone, 1984; Hill, 1989), and a more recent meta-analysis of 40 studies examining negative advertisements identified 33 studies that confirm a backlash effect (Lau & Rovner, 2009). This effect has also been found both within the United States as well as in countries with multi-party systems (Roy & Alcantara, 2016; Walter & van der Eijk, 2019). The extend to which backlash exists is also dependent on voter (and candidate) characteristics, for example, Krupnikov and Piston, 2015 demonstrate that backlash towards African-American attack sponsors is greater among white voters.

So how do parties decide to ‘go negative’ and when? The literature on this question typically formulates a rational choice explanation making the decision to start mudslinging a careful consideration of pros and cons

(Riker, 1991; Riker et al., 1996). If the benefits outweigh the costs, negative campaigning becomes a feasible strategy (Haselmayer, 2019). Accordingly, negative campaigning is a double-edged sword with potential positive and negative effects for the campaign itself.

Incivility and Toxicity

While negativity in political campaigns may take many forms, in this study we focus on incivility, and more precisely on toxicity for several reasons. First, incivility and toxicity as concepts (although not the same) have received considerable attention in computational social science (Coe et al., 2014; Hopp et al., 2020; Pascual-Ferrá et al., 2021; Theocharis et al., 2020; Vargo & Hopp, 2020). As a result, acceptable and at the same time scalable instruments to measure their prevalence in large volumes of data have been established within the community. Second, is that given one of toxicity's definitions, i.e. language that is "[...] rude, disrespectful, or unreasonable [...]" that is likely to make you leave a discussion"[...]" (Perspective API, 2021), we expect the strongest backlash if toxic ads were targeted at people who dislike this type of language, as well as the most "benefit" when toxicity is targeted at the "right" audience who are more susceptible to this kind of messaging (Nai & Maier, 2020).

It is worth discussing the nuanced differences between political incivility and toxicity. In (computational) communication literature the Perspective API, a machine learning model by Google that predicts 'toxicity' in text, has been used for annotations of political incivility (Hopp et al., 2020; Theocharis et al., 2020; Vargo & Hopp, 2020). However, the concept of incivility is not fully synonymous with toxicity. Definitions of incivility that focus on the language and tone rather than substance or the source of incivility has invited some debate in the literature (Muddiman, 2017; Rossini, 2020). For example Rossini, 2021 distinguishes between interpersonal-directed and elite-directed incivility and finds that the latter is positively associated with justified opinion expression, whereas the former is negatively associated with it (see also: Rossini, 2020). In this sense, incivility can also be a form of dissent that may even advocate for a pro-democratic stance (Edyvane, 2020). As such, political incivility is not just inflammatory language: it matters who uses it, citizens or political elites, and what an uncivil message argues *for* (Gervais, 2019). While there is considerable overlap between incivility and toxicity, the measure provided by the Perspective API focuses more on the dimension of language and tone rather than what is being argued for (or against) and the identity of the source, which are usually part of nuanced conceptual definitions for incivility (Coe et al., 2014; Muddiman, 2017; Rossini, 2020, 2021). This makes toxicity more of a sub-concept of incivility.

Negative campaigning with toxic messages is a strategy that may deliver on two particular campaign goals: the persuasion of undecided voters, and mobilization of supporters (Rohrschneider, 2002; Stuckelberger, 2019). Undecided voters may be persuaded when the opposition is successfully discredited by the negative portrayal of them while at the same time toxic ads could also galvanize the supporter base against a common threat. In accordance with *negativity bias*, it has been shown that messaging incivility can increase the memorability of the topics discussed (Sydnor, 2019) as well as encourage greater interest in voting and in politics in general (Herbst, 2010). It's not unreasonable to assume that toxicity in campaigning would have similar consequences.

Going Micro to go Negative

Democratic theory distinguishes between political candidates on the one hand, and citizens on the other. Political candidates act with the intention to maximize the likelihood of getting elected, while citizens act to increase the likelihood that they vote for someone whose policy is best in line with their interests (Downs, 1957). A political campaign plays a crucial role, as it provides citizens with information that they can use as input for deliberation. Downs (1957) perceives the citizen as a rational actor, and this perspective has been criticized because citizens' deliberative process is not always strictly rational (Budge & Farlie, 1977; Robertson, 1976). Emotional information, for example, can also play an important role (Susser et al., 2019).

A challenge for political campaigns has been that individual people respond differently to information. Some respond better to emotional information, others prefer rational information (see differential susceptibility to media effects model; Valkenburg and Peter, 2013). In other words, people are differentially susceptible to information, but political advertisers are challenged with finding out which citizen prefer what type of information and how to reach the particular subgroup with this kind of information. For example, communicating emotional messages to a large group of people comes with the risk of inducing a backlash in those people who do not prefer emotional information and would have preferred to receive factual information.

Political actors who want to make the decision to run ads that include toxic language are faced with a dilemma. While they want to reap the benefits of using the supposed (de-)mobilization effects of negative campaigning, they also don't want to suffer the consequences of backlash effects. When 'going negative' campaigns make themselves vulnerable, especially when they use toxic language and cast a wide net by showing such ads to everyone. The

more broadly targeted a toxic campaign message is, the higher is the risk of a mismatch between voter susceptibility and the toxic campaign message. Viewed this way, placing toxic ads is a high-risk, high-reward strategy that needs careful consideration. To solve this dilemma of toxic messaging, we argue that advertisers may attempt to use one of two strategies, or both:

1. narrowly target toxic messages decrease the risk of exposure to individuals that are less likely to show backlash effects (Fowler et al., 2021);
2. outsource toxic campaign messages to third party actors who are unaffiliated with the campaign, so-called *outside groups* (Chand, 2017).

Advertisers seem to be well aware of the potential backlash effects and take advantage of the situation when they do not have to fear it: outside groups are more likely to use negative ads, especially when they conceal their donor information (Chand, 2017). Given this awareness, targeting could be used as an effective strategy to mitigate potential backlash effects. By using the targeting options provided by the Facebook ad manager, advertisers may focus on specific socio-demographic groups (based on gender, age, or education level) or they might make use of so-called custom audiences. The latter provides advertisers with an especially useful tool for spreading toxic campaign messages, as advertisers have access to lists of custom audiences which they can match with additional data that reveal the susceptibility to toxicity. For example, Nai and Maier, 2020 show that incivility works in particular on people high in psychopathy traits and those who are low on agreeableness and conflict avoidance. If political advertisers get a hold of voter data that contains such information (for example through

third-party data brokers) they could upload that to Facebook and target these groups in particular with toxic messages. Of course, advertisers could choose to completely refrain from negativity and simply choose to target positive messages towards citizen to avoid backlash risks altogether. But given that strongly negative campaigning promise more effectiveness than positive messages in certain instances and social media companies afford them the possibility to manage the risk of mistargeting, the strategy of narrow targeting toxicity becomes more feasible for them.

At this point in time, there are limited and mixed findings in the literature about the usage of targeting to avoid potential mismatches between negative messages and the receiver. Analyzing the 2004 and 2008 U.S. presidential elections, Roberts, 2013 finds that campaign videos posted only online were more likely to include attacks than TV ads, arguing that online ads allow for more specific targeting than TV broadcasts. On the other hand, Fowler et al., 2021 analyze TV and Facebook advertisements but find the latter to

be more positive in tone. So, despite the fact that Facebook allows for more fine-grained targeting, advertisers seem to not have used this affordance to target more negative messages towards specific individuals. To explain this, Fowler et al., 2021 argue that campaigns are more likely to have lists of supporters which they use to ingest Facebook with target audiences and negative advertisements would be less likely to be targeted towards supporters. This account is more consistent with negative advertising being directed at potential supporters of the opponent in order to demobilize them (Ansolabehere et al., 1994; Krupnikov, 2011). However, it's important to note that Fowler et al., 2021 only compare TV with Facebook ads and do not differentiate between levels of targeting, and it may still hold true that more narrowly targeted ads on Facebook are more toxic than ads targeted at a more general audience, as we have argued. Finally, López Ortega, 2021 analyze political advertisements in Austria, Italy, Germany and Sweden and find that microtargeted ads are no more likely to be negative compared to when they are less targeted. This is contrary to our expectations but more evidence in regards to this suggested relationship is needed.

Nonetheless, four reasons can be formulated that support the idea of narrowtargeting toxic campaign messages: first, through reducing the audience size, toxic messages can be targeted towards individuals who are most susceptible to them and accepting of them. Second, the outgroup-bias of a more homogeneous group can be exploited by attacking the perceived outsiders of that group (Frederic & Falomir-Pichastor, 2018). For example, male-majority groups could be targeted with anti-immigrant attitudes using specific messages that evoke a 'hostile takeover' and a feeling of threat by incoming immigrants. Third, an adverse demobilisation of one's own supporters (or mobilisation for the opponent) can be caused by a failure to match toxic ads to a recipient who is susceptible to this kind of message (Hersh & Schaffner, 2013). Fourth, the chance of detection by the public at large is smaller when an ad is narrowly targeted because a smaller and more specific group of people is alerted to the existence of the ad.¹ Negative campaigning and political incivility in general is seen as undesirable and harmful by the public so targeting toxicity towards a smaller group of people reduces the reputational problem that such a campaigning style might cause (Krupnikov, 2015; Mutz & Reeves, 2005). And in case an ad is mistargeted due to chance, a smaller range of individuals are affected for every such message, making the risk of backlash more manageable.

Even in the case of Donald Trump's campaign style, notoriously uncivil, there is evidence that suggests his voters do negatively evaluate him for his uncivil language, although this effect does not hold with his more "die-hard

supporters” (Frimer & Skitka, 2018). This further supports the idea that more narrowly targeted audiences are likely to receive toxic campaign messages.

These considerations lead us to our first set of hypotheses, the expectation that political ads with toxic campaign messages are targeted at narrower audiences to reduce mismatch between message and receiver.

H1a: The more toxic the message of an ad, the more likely is it narrowly targeted.

The risk of backlash is also expected to vary across advertiser *types*. Parties and official campaigns who are up for election can expect a stronger backlash, because they are judged by higher standards compared to so-called “outside groups”. The term outside groups is commonly used to describe entities that spend money during electoral campaigns “independently of, and not coordinated with, candidates’ committees” (CRP, 2022; Ridout et al., 2015). This includes for example super PACs and 501(c) non-profit organizations, often nicknamed “dark money” groups, because the source of the funding is harder to track and less restrictions to spending apply. Such outside groups may have generic names that do not imply any specific affiliation with a specific candidate or ideology. For example, the “Winning Our Future” super PAC in the 2012 US election does not imply any specific affiliation but it supported Newt Gingrich as the Republican presidential candidate and heavily attacked Mitt Romney, the presidential front runner for the Republican nomination at the time (Los Angeles Times, 2012). Such outside groups are less likely to care about backlash because the source of the attack is not directly identified with the candidate standing to benefit from the attack and voters may not know who to think less off for engaging in toxicity. However, voters may still identify them as partisan or associate them with certain candidates regardless if they conceal their affiliations. Therefore, we expect outside groups to also make use of targeting to deliver toxic messages but the suggested relationship should be less strong for them than for political campaigns because they need to do less risk management:

H1b: Outside groups are more likely to narrowly target toxicity but to a lesser extent compared to official political campaigns.

As previously stated, outside groups are more likely to use negative ads, especially when they conceal their donor information (Chand, 2017). While official coordination between outside groups and official campaigns is prohibited, many instances of coordination have been documented, for example, by public

watch dogs like *Coordination Watch* (Coordination Watch, 2022; Scala, 2014). Coordination between outside groups and official campaigns typically occurs because campaigns want to make use of additional funds they wouldn't legally be allowed to spend themselves. We theorize that an additional potential benefit that campaigns may derive from coordination is to outsource toxic campaigning without risking backlash. Given that outside groups are not affiliated with candidates or parties officially, they may do the "dirty work" for them by placing the most toxic campaign messages while official campaigns can run more clean and civil ads. It is also the case that no explicit coordination between outside groups and official political campaigns is necessary for this outcome to happen. There might also just be an implicit understanding that outside groups can take part in more toxic advertising whereas official campaigns need to be more careful. The following hypothesis is thus formulated:

H2: Outside groups are more likely to place toxic ads than official political campaigns.

Data

We retrieved all online advertisements of the 2020 US election that were placed on Facebook platforms (incl. Instagram) between April 3rd and November 3rd 2020 via the Facebook Ad library API (Constine, 2019). In total, we were able to collect a data set of 3.64 million ads that were placed in that time period. However, closer inspection revealed that many of the advertisers and their advertisements are not political. Unfortunately, as the Ad library does not provide any information about the advertisers themselves, this makes it difficult to narrow down the data set to a relevant political sample for analysis. In order to learn more about the sponsors of the ads, we used various data sets and merged them with the collected Facebook Ad library data:

- Federal Election Commission (FEC, 2021)
 - Committees registered with the FEC in the 2020 election cycle (18,295 committees, Super PACs, Corporations, Non-Profits etc.)
- Center for Responsive Politics (CRP, 2021b)
 - Trump Facebook Political Advertisers (181 distinct funding disclaimers)
 - Biden Facebook Political Advertisers (68 distinct funding disclaimers)
 - "Dark Money" groups (1,064 distinct groups) (CRP, 2021a)

After matching the FEC and CRP data with the Facebook Ad library data via the disclosed ‘funding entity’ disclaimers less than half (1.42m) of all ads remain. Manually checking the ads of the top 20 occurring funding entities that we were unable to match (596,899 ads or 28% of all unmatched ads), reveals that they are mostly non-political advertisers. Besides some apolitical commercial advertisers, the non-matched accounts belong to local news aggregators and polling companies recruiting participants for surveys. 257k ads just lack a funding entity entirely, but further investigating the top 20 occurring advertiser pages here (84,202 ads, or 32.73% of all ads without funding entities) reveals that they are exclusively non-political as well. It is important to note that the matching procedure with publicly available sources such as the FEC means that our sample is likely to be biased towards advertisers that are not trying to obfuscate their identity, since they are using consistent naming in both their FEC files and their Facebook disclaimers. Using the CRP data, especially the “Dark Money” groups list, should partly alleviate this problem but we cannot account for advertisers that simply use different disclaimers for their ads.

The next step is to narrow down the data set to a relevant period. In this case we chose the time period of *3 months* before election day between August 3rd and November 3rd 2020 which coincides closely with the DNC (17. Aug. – 20. Aug. 2020) and RNC (24. Aug. – 27. Aug. 2020) conventions which officially named Trump and Biden as the nominees of their respective parties and marks the points where a lot of fundraising money started flowing to their campaigns. After limiting the data to just the 3 months before election day we are left with a sample of 946,820 ads. However, narrowing down the data set in this way may affect the results in itself and it is thus good to be aware of “researcher degree of freedom” (Wicherts et al., 2016). In order to address this we perform *multiverse analyses* which seeks to account for all alternative coding and data processing decisions that researchers could have made and checks whether results are dependent on them (Simonsohn et al., 2015; Steegen et al., 2016). In the following we briefly discuss the variables used throughout the various models that we construct for the analysis.

Advertiser Types

Using the FEC and CRP data, we grouped the advertisers into *Political Campaigns* and *Outside groups* as follows:

- Political Campaigns:
 - House, Senate, or Presidential campaigns
 - Parties: Democratic or Republican
 - Joint Fundraising Committees: Democrat and Republican state parties joint committee with presidential candidates

- Outside groups:
 - Hybrid PACs
 - (Super) PACs
 - 501(c) Organizations (non-profits)
 - Corporations
 - Political retailers
 - Fan pages

For the purpose of this study, we define an entity as an outside group if it runs political ads and is not officially affiliated with any party or candidate. The FEC marks advertisers as “independent-expenditure” which makes it easy for us to identify those advertisers that should be acting independently of official candidate and party campaigns. This includes for example the “America First Action” Super PAC which supported Donald Trump for president and the progressive “MOVEON.ORG POLITICAL ACTION” hybrid PAC. Further, CRP identifies “dark money” groups which were all marked as outside groups in this analysis, which includes corporations, companies and also non-profits such as the conservative “Turning Point USA” organization.

The data also includes a series of non-text ads such as videos and images, however the media itself is not provided by the Facebook Ad library API and they have to be retroactively scraped. Given that video and images are likely to include a range of information relevant for measuring the incivility of an ad, we employ various computational methods to gather as much text as possible for any given ad: 1.) for videos, we first extract the audio and then apply Mozilla DeepSpeech to transcribe the speech used in the ads (Hannun et al., 2014), 2.) for images, we use Google Cloud Vision API to apply Optical Character Recognition (OCR) (Vaithyanathan & Muniraj, 2019). In the process of downloading and scoring the text data a small fraction of ads were lost (due to inability to download the ad media) and the final data set consists of 912,110 ads.

Unfortunately, the Facebook ad library does not provide specific information about targeting criteria. In absence of transparency of targeting criteria used by advertisers on Facebook platforms, we employed the *potential reach* variable provided in the ad library as a rough measure of targeting.

Targeting

According to Facebook, “[Potential Reach] estimates how many people your ad could potentially reach depending on the targeting and ad placement options you select while creating an ad” (Facebook Business Help Center, 2021).² The metric is always calculated before an advertisement is placed

and directly follows from the strategic boundaries that are formulated by the advertiser. Based on this definition, a higher potential reach means that an advertiser was trying to target a broader range of citizen, while a lower potential reach means the advertiser was engaging in narrower targeting by, for example, stacking multiple exclusive targeting criteria.³ Potential reach takes the following values, 100 – 1.000 individuals reached, 1.001 – 5.000, 5.001 – 10.000, 10.001 – 50.000, 50.001 – 100.000, 100.001 – 500.000, 500.001 – 1 million and above 1 million.

Toxicity

The toxicity in ads can be detected by using Google's Perspective API that gives access to a machine learning model that scores "toxicity" (Hopp et al., 2020). At its heart, Perspective is a supervised machine learning algorithm that is trained on millions of annotations by tens of thousands of individuals, ranging from crowd staff to New York Times comment moderators (Wakabayashi, 2017). The Perspective API provides a measure that ranges from 0 to 1, indicating the probability of whether a given piece of text is toxic or not. In order to retrieve the toxicity scores we use the R wrapper for the API called *peRerspective* (Votta, 2021). For each ad, we run both the accompanying description of the ad and also the extracted text from image and video through the Perspective API. Prior research by Hopp et al., 2020 shows that the API components provide generally reliable measures of uncivil political discourse on Facebook and Twitter and Vargo and Hopp, 2020 have shown that it works well within the context of political advertisements.

Control Variables

We control the models for audience shares, i.e. the distribution of audiences who saw the ads by age, gender and state in percent. For example, an advertisement may have only be seen by men but also by both male and female audiences. This means, an ad might be targeted at people of all ages and/or all genders or they might for example be specifically targeted at younger people (and exclude middle-aged and older audiences, male or females). Further, we control for the "maximum number of reached people", i.e. the calculated number of Facebook users an ad could have reached as indicated by the audience share metrics (see appendix for more detailed calculation). Our main dependent variable for H₁ is particularly focused on the audience size, which is bound to differ systematically by the size of the campaign. Therefore 'maximum reach' is particularly important to include as a control variable because it allows us to compare between different campaign sizes. Advertisers may have only ever intended to target

a specific state because a candidate was only electable in that state (e.g. a Senate race) so taking into account the maximum number of people the ad could have reached controls for that fact.

We further control the models for ad run time (how *long* the ad ran) and unique ad runs *how often the ad ran*. We could expect that audience shares, ad run time, unique ad runs, as well as the timing and the money spend on the ad might be related to toxicity because toxic ads could run shorter, less often and to specific demographic groups. This relation is grounded in the idea that toxic ads come with risk for the advertiser. A longer ad run time, as well as ads that ran more often, may increase the risk of exposing the ad to more people who find the ad objectionable, for instance. We also control for the timing of the ad (how close to the election the ad ran), whether the advertiser was pro-Republican or pro-Democrat party and how much money was spend on the ad. A toxic ad that ran closer to election day might be more of a risk than one that is far out, and there might be a significant difference in both targeting and toxic messaging strategy depending on the political position of the advertiser.

A more detailed explanation of used control variables can be found in the appendix.

Analysis

Before we attempt to test the hypotheses it is worth exploring some basic descriptives of the data. Let's first take a look at the dependent variable: *Targeting*. As previously stated the variable comes in 8 distinct targeting ranges that (somewhat arbitrarily) vary in range sizes (a difference of 900 for the smallest and 500k for the biggest range). Figure 1 shows that roughly a third of ads are targeted at a broad audience, albeit a majority of ads in our sample are targeted at audience sizes smaller than a million users (65.20%). The tendency to target groups smaller than a million does not differ drastically between official campaigns (which include political candidates and parties) and outside groups (64.88% vs. 66.43%). However, the comparison changes when we set different cut-off values: for example 29.34% of outside group ads are targeted at audience sizes smaller than 50k users, whereas only 18.70% of political campaign ads are targeted at an audience that small. Next, we examine our main independent variable: *Toxicity*. Figure 2 shows the distribution of the variable and some descriptive statistics. Outside groups have a slightly lower mean probability score of toxicity (mean = 0.17) than campaigns (mean = 0.20).

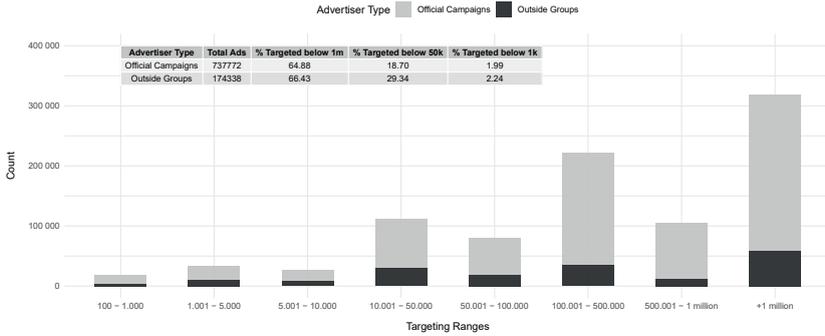


Figure 1. Targeting - Descriptive Statistics

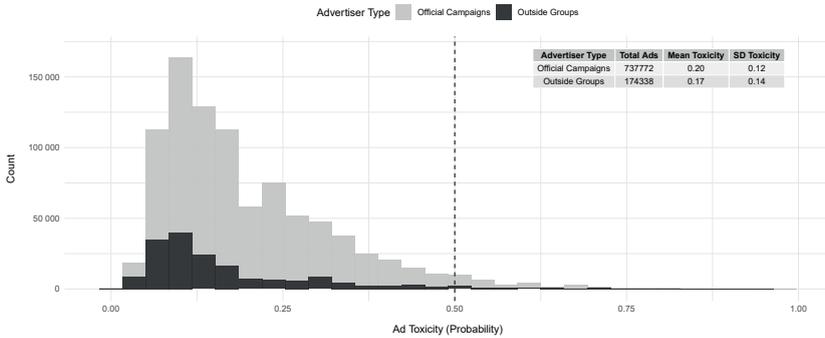


Figure 2. Ad Toxicity - Descriptive Statistics

Analysis: H1

Hypothesis H1a is investigated using multilevel ordinal logistic regressions. A multilevel model is applied because advertisements are nested within advertisers and an advertiser may run thousands of ads. To account for this, we estimate a random intercept for each page that places an ad. The dependent variable is an ordinal variable of targeting, ordered in the direction of *smaller audience size* (i.e. first value is an audience size of +1 million and the last value is an audience size of 100 – 1.000) and toxicity as main independent variable. A null model fitted to the data suggests that 48.4% of the variance of targeting are found on the

page-level, well justifying the use of a multilevel infrastructure (ICC = 0.484). Model 1 includes all control variables, Model 2 adds variables of interest, i.e. toxicity and advertiser type (Political campaign vs. outside groups) and Model 3 adds an interaction between the two. The results are

Table 1. Odds Ratios for Model 1 – 3 DV: Targeting (ordinal)

	Model 1	Model 2	Model 3
Political Affiliation (ref. Pro-Rep)	0.85 (0.10)	0.72** (0.08)	0.72** (0.08)
Age Group: 13-34 (ref. all ages)	0.81*** (0.02)	0.82*** (0.02)	0.82*** (0.02)
Age Group: 44-54	0.67*** (0.04)	0.67*** (0.04)	0.67*** (0.04)
Age Group: 55+	0.97 (0.02)	0.97 (0.02)	0.97 (0.02)
Age Group: Other	0.78*** (0.00)	0.78*** (0.00)	0.78*** (0.00)
Gender: male (ref. male and female)	0.88*** (0.02)	0.88*** (0.02)	0.88*** (0.02)
Gender: female	0.69*** (0.01)	0.70*** (0.01)	0.70*** (0.01)
log. Max Reached	0.70*** (0.00)	0.70*** (0.00)	0.70*** (0.00)
log. Ad runs	0.89*** (0.00)	0.89*** (0.00)	0.89*** (0.00)
log. Ad run time	1.07*** (0.00)	1.07*** (0.00)	1.07*** (0.00)
log. Closeness to Election	1.17*** (0.00)	1.17*** (0.00)	1.17*** (0.00)
log. Spending (lower bound)	0.96*** (0.00)	0.96*** (0.00)	0.96*** (0.00)
Toxicity		1.05 (0.03)	1.22*** (0.04)
Advertiser Type: Outside groups (ref. Pol. Camp.)		0.27*** (0.03)	0.29*** (0.03)
Toxicity × Advertiser Type: Outside groups			0.60*** (0.04)
AIC	-239900.61	-239843.84	-239839.10
Log Likelihood	-239880.61	-239821.84	-239816.10
Deviance	479761.22	479643.68	479632.21
Num. obs.	912110	912110	912110
Num. groups: Advertiser (pages)	1663	1663	1663

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Coefficients are odds ratios.

shown in Table 1. First, we compare whether adding our main variables of interest significantly increases model fit by comparing Model 1 and Model 2: a statistically significant likelihood ratio test between Model 1 and 2 reveals that Model 2 has a better fit than Model 1 ($\chi^2(21) = 117.54$, $p < 0.001$). In Model 2 we can observe that there is no statistically significant relationship between

toxicity and the narrowtargeting of an ad (Model 2: OR=1.05, 95%CI [1.00, 1.11]), which is not in line with our expectations in H1a that suggested ads with higher toxicity to have smaller audience sizes.

To test the moderator hypothesis H1b, we estimate an interaction between toxicity and advertiser type in Model 3. A statistically significant likelihood ratio test between Model 3 and 2 reveals that Model 3 has a better fit than Model 2 ($\chi^2(22) = 11.47, p < 0.001$). While we do not observe a statistically significant association between toxicity and targeting in Model 2, it is possible that a cross-interaction masks this relationship. Model 3 (Table 1) indeed finds such a cross-interaction: the toxicity coefficient turns positive and statistically significant (Model 3: OR=1.22, 95%CI [1.14, 1.30]) when including the interaction. The interaction between advertiser type and toxicity is also statistically significant (Model 3: OR=0.60, 95%CI [0.53, 0.67]) which, given an odds ratio coefficient of below 1, suggests that outside groups are less likely to narrowly target their toxicity than official political campaigns, in line with expectations in H1b.

Figure 3 zooms in on the significant interaction in Model 3 between the advertiser type (Political Campaigns and Outside groups) and toxicity and further illuminates the missing statistical significance for toxicity alone in Model 2. Figure 3 shows the toxicity on the x-axis and predicted probabilities for each targeting range on the y-axis. It is divided in 8 panels, each for one of the 8 audience sizes of the targeting variable. Further, advertisements coming from an official campaign are shown with *solid* lines and those from outside groups are *dotted*.

Focusing on the upper right panel of Figure 3, we observe that the probability for an ad to be broadly targeted (an audience size of +1 million) increases with higher levels of toxicity for outside groups, whereas it decreases for official campaigns. As we go further along the panels from left to right and from top to bottom, i.e. to smaller audience sizes, we can observe that official campaigns and outside groups switch places. For example, for very small audience sizes (1000 to 5000 and 100 to 1000), we observe that the probability for official campaign ads to be targeted to such small audiences increases with higher toxicity, whereas it slightly decreases for outside groups. This is close to what we expected but differs from our expectations in a key aspect. We hypothesized that both outside groups and official campaigns target their more toxic ads to smaller audiences, however, the results suggest outside groups are more likely to target greater audience sizes with toxicity (top-left panel).

It is worth investigating how this relationship differs depending on different time cut-off points relative to the election. The left panel of Figure A1

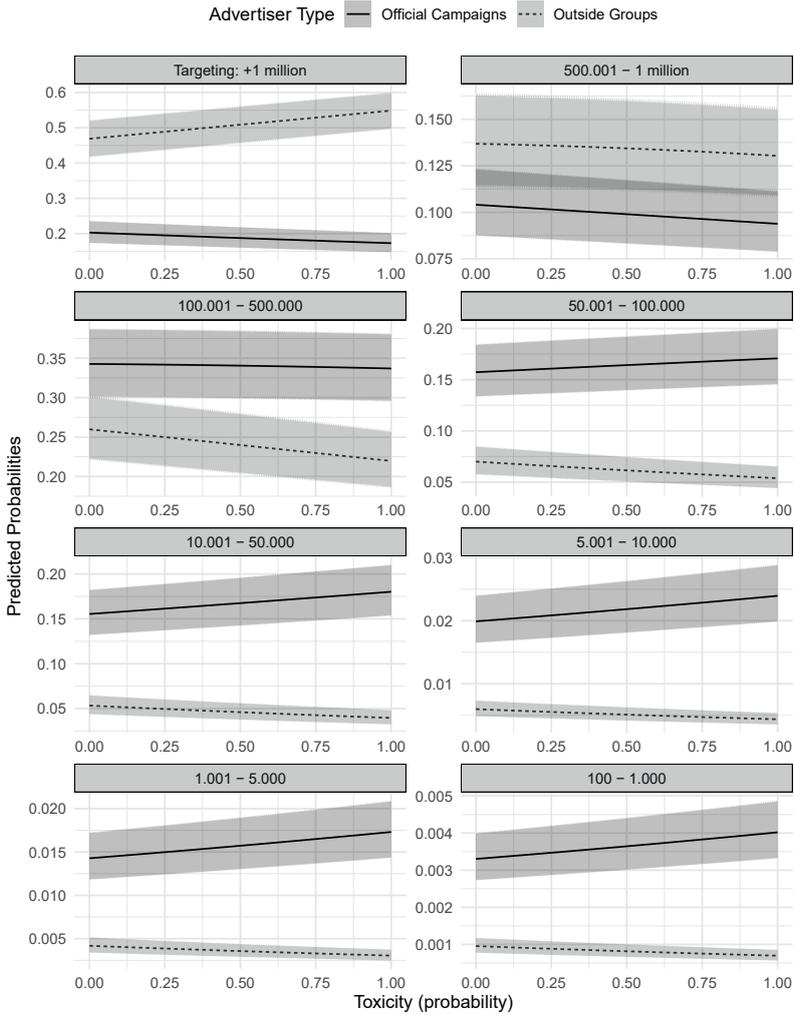


Figure 3. Model 3: Predicted Probabilities of Toxicity X Advertiser Type

(in the Appendix) shows that adding the advertiser type (official political campaigns vs. outside groups) interaction with toxicity only makes a difference 4 months before election day. When including data for 5 months before the election and later, toxicity is more likely to be directed at bigger audience sizes.

Accordingly, we can say that H1a only holds when we consider the source of the advertiser type and advertisements at maximum 4 months before the

election, whereas we find the opposite of what we expected when considering advertisements that were placed 5 months before the election and later. Nonetheless, the right panel of Figure A1 shows consistently that outside groups are more likely to target toxicity at greater audience sizes compared to official political campaigns, whether we consider advertisements 7 months out or just 1 month before the election. To recall, H1b expected that the relationship proposed in H1a (toxic messages being more narrowly targeted) is smaller for outside groups than for official political campaigns, yet we find that as the toxicity of an ad increases it is more likely to be targeted at broad audiences. This leads us to conclude that there is only mixed evidence for H1b.

Analysis: H2

Next, we investigate hypothesis H2 which expects that outside groups are more likely to be toxic than political campaigns. We test this hypothesis by using multilevel linear regression with toxicity as dependent variable and the advertiser type (political campaigns vs. outside groups) as main independent variable. As before, we estimate a random intercept for each page that placed an advertisement. A null model fitted to the data suggests that 28.3% of the variance of toxicity are found on the page-level, well justifying the use of a multilevel infrastructure (ICC = 0.283). The following two multilevel models are estimated: Model 4 includes various control variables that were introduced before and Model 5 introduces the main independent variable advertiser type (reference category: official political campaigns). The results in Table 2 show that ads placed by outside groups are more likely to be toxic compared to official political campaigns (beta = 0.01, $p < 0.001$; std. beta = 0.11, 95% CI [0.05, 0.17]), which supports our expectations in H2. A significant likelihood ratio test between Model 4 and 5 reveals that Model 5 has a better fit than Model 4 ($\chi^2(1) = 11.877$, $p < 0.001$). Finally, testing for different time cut-off points (Figure A5 in appendix) reveals that this result is robust to any specification regardless of whether we cut-off the data any other month before the election. The estimated effects in Model 4 and 5, as well as the robustness test, all find support for hypothesis H2: outside groups are more likely to place toxic ads than official political campaigns.

Conclusions

Our exploratory study of how toxicity is targeted during the 2020 US election has revealed several noteworthy findings. First, we argued that political advertisers want to make use of toxicity in campaigning

Table 2. Coefficients for Model 4 – 5. DV: Toxicity

	Model 4	Model 5
Intercept	–0.35*** (0.02)	–0.39*** (0.02)
Political Affiliation (Pro-Dem/Pro-Rep)	0.25*** (0.03)	0.27*** (0.03)
Age Group: 13-34 (ref. all ages)	0.07*** (0.01)	0.07*** (0.01)
Age Group: 44-54	–0.09*** (0.01)	–0.09*** (0.01)
Age Group: 55+	0.02** (0.01)	0.02** (0.01)
Age Group: Other	0.01** (0.00)	0.01** (0.00)
Gender: female (ref. male and female)	–0.01** (0.00)	–0.01** (0.00)
Gender: male	0.01 (0.01)	0.01 (0.01)
log. Max Reached	0.00 (0.00)	–0.00 (0.00)
log. Ad runs	0.06*** (0.00)	0.06*** (0.00)
log. Ad run time	0.00 (0.00)	0.00 (0.00)
log. Closeness to Election	0.06*** (0.00)	0.06*** (0.00)
log. Spending	0.02*** (0.00)	0.02*** (0.00)
Advertiser Type: Outside groups (ref. Pol. Camp.)		0.11*** (0.03)
AIC	–1416431.48	–1416432.14
BIC	–1416255.63	–1416244.56
Log Likelihood	708230.74	708232.07
Num. obs.	912110	912110
Num. groups: Advertiser (pages)	1663	1663

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Coefficients are unstandardized.

because of so-called *negativity bias*, the human tendency to assign more value to negative-valenced information over neutral- or positive-valenced information (Fiske, 1980; Rozin & Royzman, 2001). Incivility in particular has been shown to increase the memorability of the topics discussed (Sydnor, 2019) as well as encourage greater interest in voting and in politics in general (Herbst, 2010). However, some people strongly dislike toxic language and may punish the sponsor rather than lower the opinion

of the attack target in what is known as *backlash effect* (Garramone, 1984; Hill, 1989; Lau & Rovner, 2009). To resolve this issue we expected that toxic messages would be more likely to be targeted at narrower audiences to avoid potential mismatches between message and receiver and the backlash that can arise from that. When testing this hypothesis (H1a) we find mixed evidence in support for this idea. We find that this relationship is only found when we consider the sponsor (official political campaigns vs. outside groups) and the timing of the ad. The suggested relationship only holds for advertisements that are placed no later than 4 months before the election, whereas when considering advertisements placed 5 months and later we observe the *opposite* relationship from what we expected.

Further, in H1b we hypothesized that both official campaigns and outside groups would try to target smaller audience sizes with toxic messaging. However, outside groups were expected to target toxicity to smaller groups to a lesser extent because they need to manage the backlash risk of mismatching toxic messages less than official campaigns. The results suggest something similar but are not completely in line with expectations: we find that only toxic ads sponsored by official political campaigns are more likely to be narrowly targeted, whereas outside groups are more likely to target toxicity broadly. Nonetheless, we find that regardless of which timeframe we consider for our analysis, more toxic advertisements by outside groups are always more likely to be targeted at broader audiences than official campaigns. This leads us to conclude that there is only mixed evidence for H1b.

Finally, we expected that outside groups are more likely to be sending toxic messages than official political campaigns as the latter may try to outsource the risk of placing toxic ads to outside groups (H2). This is supported by the evidence and is in line with previous scholarship on negative campaigning and divisive ads on Facebook (Chand, 2017; Kim et al., 2018).

Let us recall the main research question asked in the beginning of this paper:

To what extent is political microtargeting used by political advertisers to deliver toxic campaign messages?

Based on the sampled data we can conclude that political microtargeting, in the form of narrowly targeting smaller audiences, was used to deliver toxic messages during the 2020 US election on Facebook and Instagram. Our

analysis however, suggests that toxicity only leads to “narrowly” targeted campaign messages under certain circumstances and there is also some evidence that toxicity was in fact more broadly targeted when it comes to dates further away from election day. Nonetheless, it is important to note that a substantial share of ads was indeed targeted at small audiences. So if the toxicity of a message does not always explain well whether it was narrowly targeted or not, it is worth thinking about what other factors might influence the decision to “go micro”. Rather than the tone of the message, another interesting factor could be to investigate the substance and the issues discussed in the advertisements.

Future research could look into whether specific topics are more likely to be targeted at niche audiences that care more about the content of the ads or whether microtargeting leads to a diversification of topics (López Ortega, 2021).

Limitations

In our study, we defined microtargeting as “narrowtargeting” and measured this concept based on the intended audience size, using Facebook’s *potential reach* measure. However, microtargeting goes far beyond just targeting a smaller audience: the practice of microtargeting refers to targeting a *specific* audience based on socio-demographics and other key factors about (inferred) identity of an user. Focusing on narrowcasting and potential reach can be justified in the context of this paper because “broadcasting” a message has a higher potential for mismatching and that is at the core of the paper’s theoretical argument. It is further the only proxy available to the researchers due to the lack of ad transparency in the Facebook ad library. However, a better measure for microtargeting would include the actual targeting criteria that advertisers select and whether custom or lookalike audiences are used, so that researchers can better assess the intended audiences. To our knowledge, only two social media platforms provide ad libraries that contain this kind of information: Snapchat and Google. Snapchat however represents a much smaller platform whereas Google affords advertisers with much more limited targeting options relative to Facebook. We endorse guidelines by ad transparency activist groups like Who Targets Me to ensure data quality and cross-platform comparability so that researchers and the public can better understand how political advertisers try to target voters during elections (Who Targets Me, 2020, 2021).

Another limitation of this research is that we are only looking at one event: the US 2020 election. Given a raging pandemic and an, arguably, extremely uncivil sitting president, the 2020 election could be considered

an outlier event and the patterns found in this election may not necessarily replicate across other elections within the US or elsewhere in the world. Future research could expand the research design here to study other countries and elections across the world. In particular, researchers have noted that political microtargeting, while originating in the United States (Kreiss, 2016), has now increasingly received more attention in European countries (Anstead, 2017; Dobber et al., 2019; Dobber et al., 2017; Kruschinski & Haller, 2017). The theoretical argument made in this paper should apply more strongly in societies where negative campaigning might be rarer than in the US context and citizens less exposed to its effects. This could mean that they are more likely to punish politicians engaging in it, implying that the proposed risk management strategy of targeting specific audiences becomes even more important in countries with lower levels of political incivility. However, one initial study on microtargeting negative messages in some European countries find no such relationship (López Ortega, 2021). More research is needed to see whether these results hold for other countries and elections.

Further, this study focused on outside groups in comparison to official political campaigns in particular. There are however many different types of outside groups (e.g. super PACs, non-profits etc.), yet the nuance between these organizations was out of the scope for this paper. Nonetheless, we encourage future research to investigate the use of microtargeting by different kinds of political advertisers and whether they make use of it for targeting negative, uncivil or toxic campaign messages differently.

Implications

Notwithstanding its limitations, this study contributes to the literature in two ways: first, it makes a theoretical contribution as it is adding to the small but emerging literature which theorizes about how inflammatory messages could be microtargeted via the sophisticated ad delivery systems afforded by social media platforms, in this case: Facebook and Instagram. We drew on democratic theory to argue that some citizens prefer different kinds of information, be it factual or loaded with emotion (Budge & Farlie, 1977; Robertson, 1976; Susser et al., 2019). Throughout this paper, we argued that political advertisers need to do risk management when they want to reach citizens with toxic messages, because some people are less persuadable by such messages in accordance with negativity bias, and more importantly, likely to punish the advertiser if they dislike this kind of campaign style.

Second, our results connect to the existing but scarce literature that empirically investigates how negative campaign messages are delivered

to users in environments that allow for very specific targeting: Fowler et al., 2021 show that ads placed by the same campaign are more positive on Facebook, where more narrowcasting is possible, than they are on TV, where broadcasting is the default. Despite the fact that Facebook would allow political advertisers to avoid backlash by targeting negative messages to specific voter groups, they do not seem to always make use of this affordance. In particular, it is worth noting that toxicity seems to be more broadly targeted for ads that are placed 5 months before the election and later, and the closer we get to election day, the higher are levels of toxicity and the more likely they are to be narrowly targeted. Scholars of negative campaigning have documented that attack ads become more prevalent closer to election day because candidates first need to establish what they stand for before they attack the other side (Damore, 2002). However, given that a significant share of people make up their minds in the final weeks of the campaign, 15% in the 2020 US election (Pew Research Center, 2020), the risk that your toxic campaign messages backfire is increased since these voters could potentially swing either way. This might explain why we only find relationships in the expected direction for dates closer to election day.

Further, we find some support that political campaigns outsource their toxicity to outside groups, as they are more likely to place uncivil ads than political campaigns. Kim et al., 2018 in particular have found evidence of outside groups engaging in divisive and inflammatory messaging via the Facebook ad delivery system during the 2016 US election. Our findings suggest that the issue of outside groups placing divisive ads has at least remained a factor in the 2020 US elections.

In this study, we proposed that microtargeting is a strategic choice that advertisers can make in order to reduce potential backlash effects. However, political campaign staff are not masterminds who can perfectly steer the right messages to the right people, even if they have access to a vast quantity of data and the sophisticated targeting choices of social media platforms under their finger tips. In fact, the entity primarily responsible for who sees which Facebook ad is not the advertiser themselves but the Facebook ad delivery algorithm which learns over time what kind of individuals are more likely to fulfill the ad goal, whether it would be making a donation, solicit voter information or signing up to volunteer. Thus, negative campaign messages may be delivered to people more susceptible to them not because that is what advertisers intentionally made happen. Rather, a fair share of the decision comes from the Facebook ad delivery algorithm which learns over time that these kind of ads attract a specific kind of audience and it automatically optimizes for that. This microtargeting without microtargeter

or “automated microtargeting” is a phenomenon that is supported by studies that find the Facebook ad delivery algorithm is steering ads towards people that may find the ads more “relevant”, as Facebook puts it (Ali et al., 2019; Ali et al., 2021). Future studies could build on these designs and try to disentangle whether ads reach their targets primarily due to human targeting choices or whether it is the algorithm that optimizes for specific goals, as likely both are involved.

To conclude, the topic of how negative campaign messages may reach voters with the help of microtargeting remains understudied. With this study we only scratched the surface of potential research. Future research could build upon our approach to study the interplay between targeting and negativity in different elections and contexts.

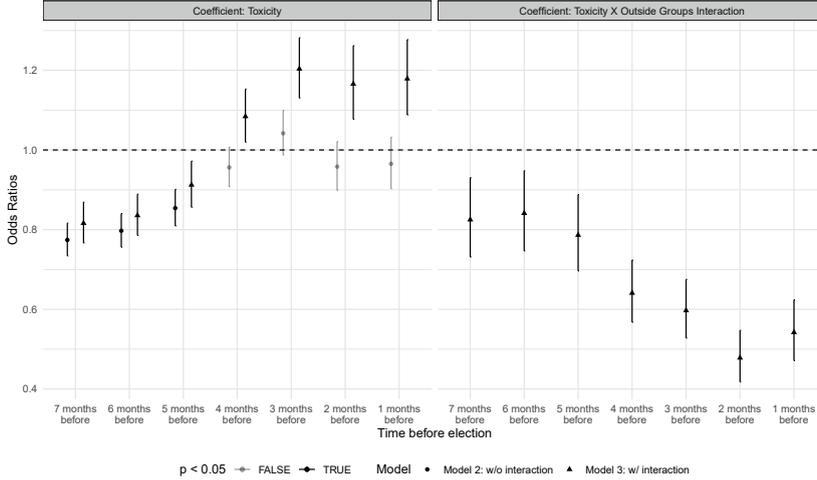
Notes

1. It is worth mentioning though that the mere existence of online ad libraries may have an effect on the strategic behaviour by campaigns because ads can now be checked retrospectively by researchers and the public. Still, platform ad libraries have their limitations and are known to miss an unknown proportion of political ads.
2. Note: potential reach was renamed by Meta to “Estimated Audience Size” in late 2021
3. It is possible of course that an ad could have relatively high potential reach but still be targeted at a very specific (social) group. An ad might for example be targeted at low-income rural people in every swing-state, which would be a big group of people but could be considered microtargeting. However, a very low potential reach would in fact imply that only a very narrow group of people was targeted whereas a very high potential reach implies a broad unfocused audience. In addition, potential reach is the only measurement in the ad library that could be considered a proxy for targeting and while it has flaws it also has utility because it is directly tied to targeting criteria.

Appendix

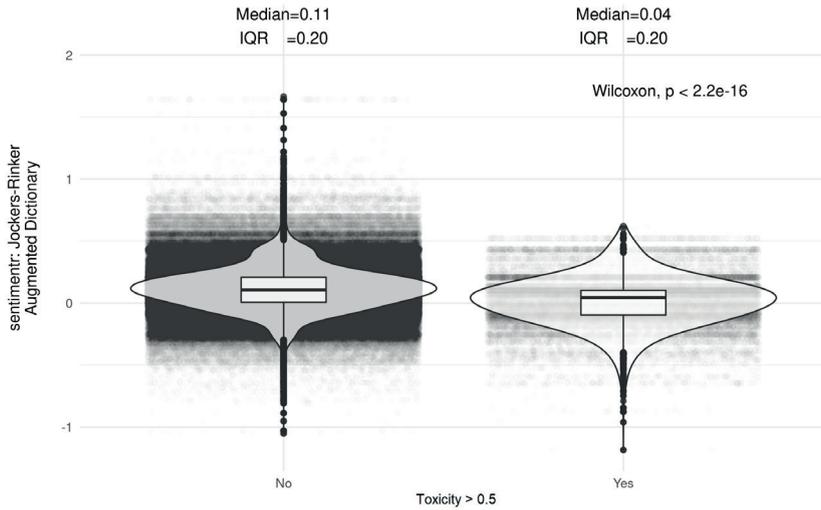
Figure A1

Odds ratios of Toxicity in Model 2 and Model 3: different time cut-offs



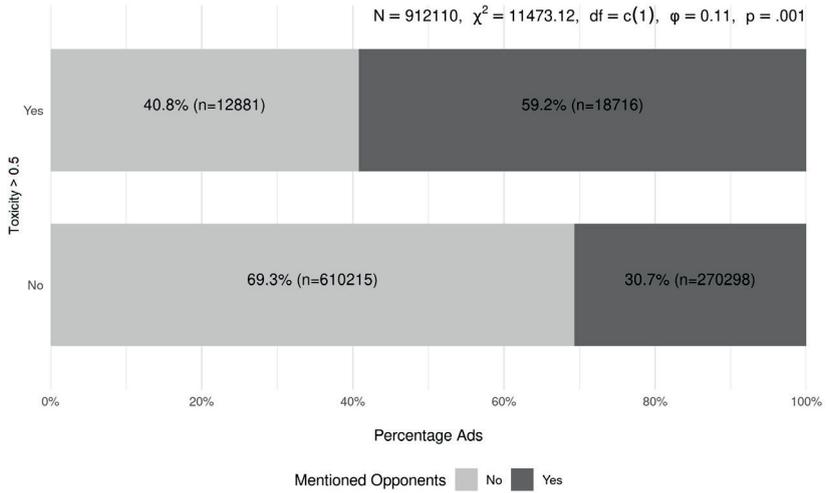
Odds ratios (and 95% confidence intervals) for Model 2 and 3 specifications varying by different time cut-off points. In the left panel, we observe that toxicity in advertisements 4 months before the election are likely to be directed at smaller groups, but only if we estimate an interaction with outside groups. In the right panel which shows an interaction between advertiser type (official political campaigns vs. outside groups) we observe consistently that outside groups are more likely to target greater audience sizes with toxicity compared to official political campaigns, whether we consider advertisements 7 months out or just 1 month before the election.

Figure A2
Sentiment and Incivility



sentimentr provides a so-called augmented dictionary which does not only match positive and negative words to assess sentiments but also integrates weighting for valence shifters, negators and amplifiers/deamplifiers, which are used to reverse, raise, and diminish the influence of a polarized word, respectively. See <https://github.com/trinker/sentimentr> for more description of sentiment analysis tool.

Figure A3
Mentioned Opponent and Toxicity



Mentioned Opponents for pro-Democrat advertisers: trump, mike pence, republican(s), conserva- tive(s). Mentioned Opponents for pro-Republican advertisers: biden, kamala harris, democrat(s), liberal(s).

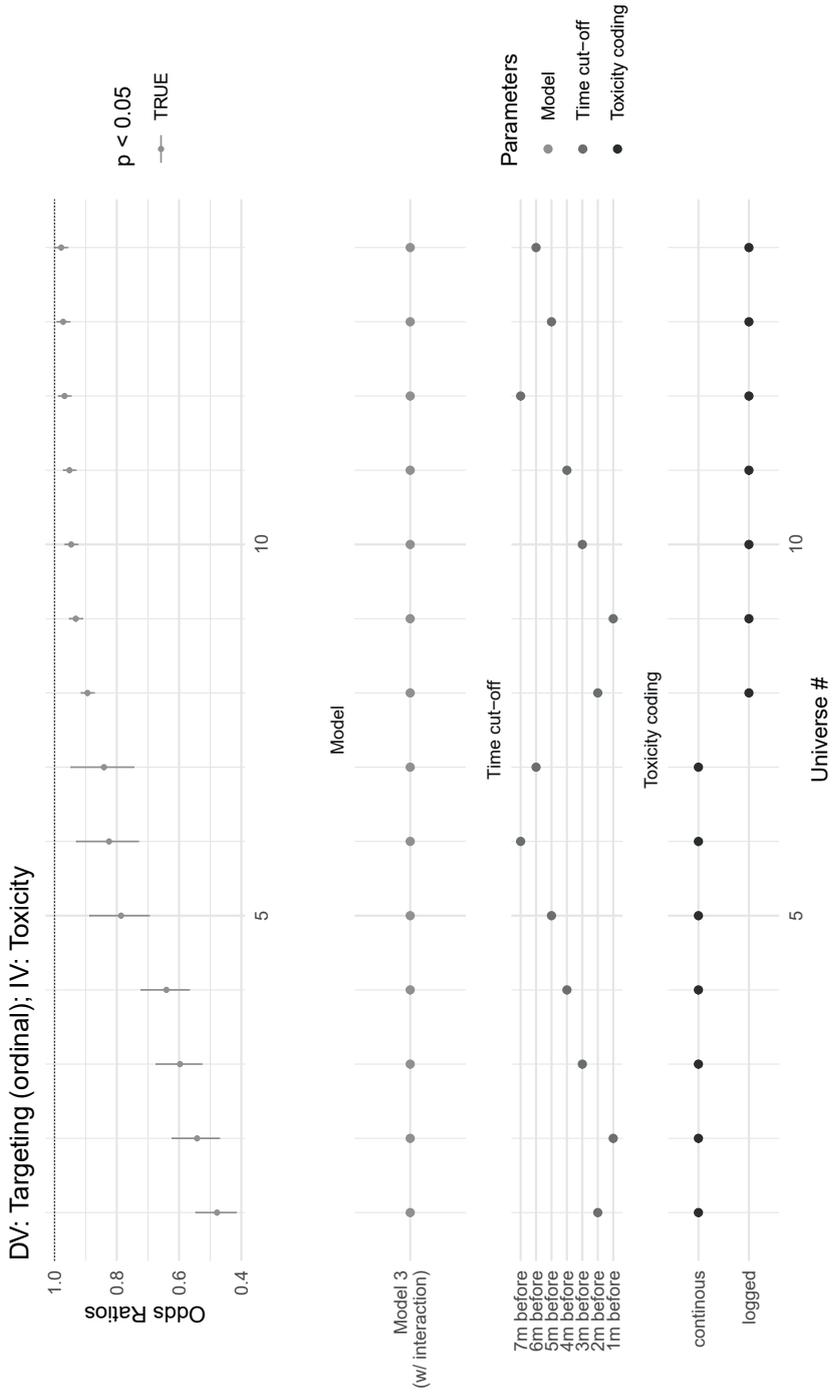
Table A1
Toxicity Scores: Examples

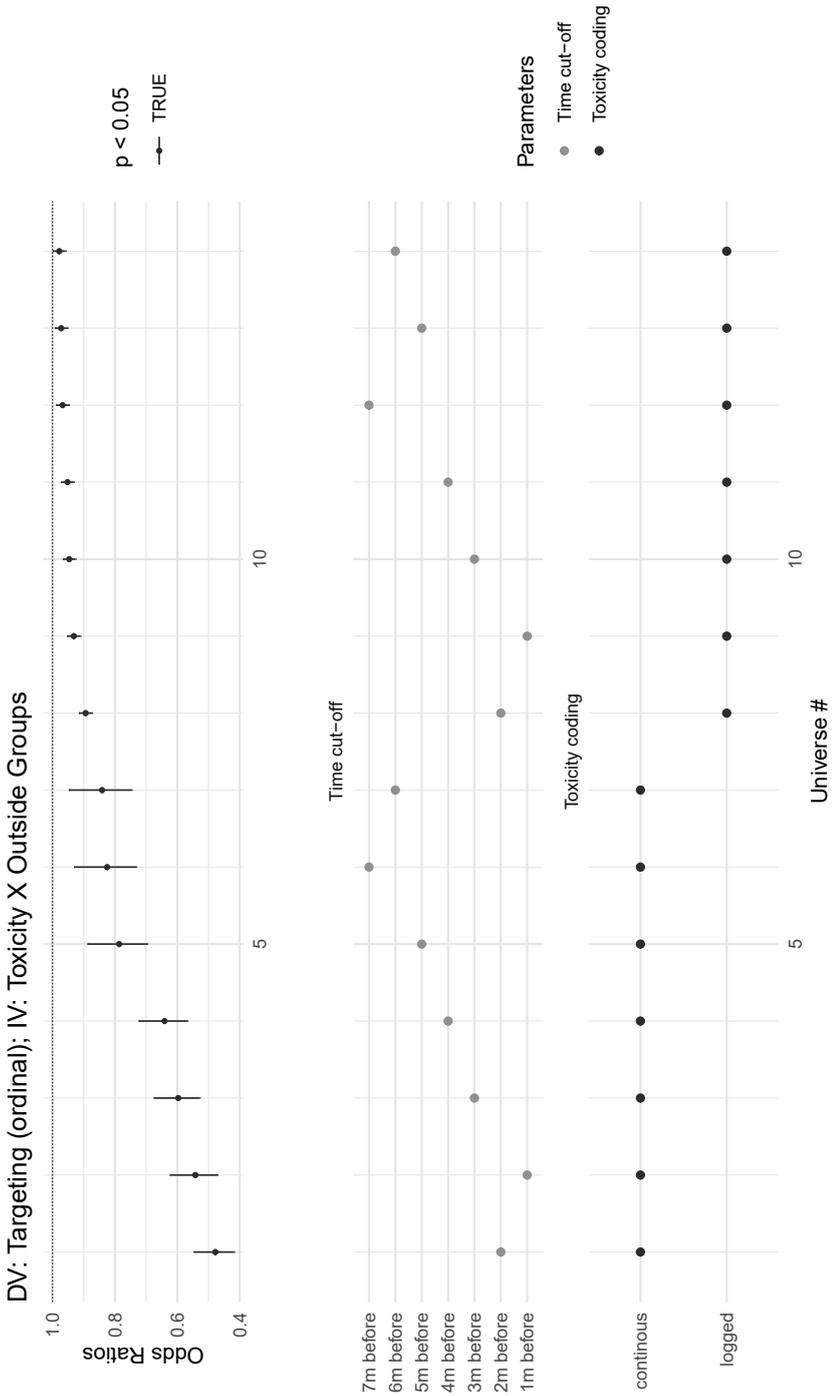
Ad ID	Page Name	TOXICITY	Toxicity > 0.5	Political Affiliation	(Extracted) Text
1000351430440885	Joe Biden	0.07	No	Pro-Dem	Biden for President United, we win Team Biden-Harris will bring America together. BIDEN HARRIS
320690692468493	Omar Navarro	0.11	No	Pro-Rep	Beverly Hills Trump Rally with ANTIFA & BLM Protesting. Make sure you support my podcast https://www.gofundme.com/f/realomarnavarro Please Follow & like my Facebook.
142427278777380	Joe Biden	0.15	No	Pro-Dem	Biden for President A clean energy revolution if we can harness all of America's energy and talents, we can turn the threat of climate change into an opportunity. Joe Biden will lead us. CLEAN ENERGY TO TO FUEL OUR ECONOMY BIDEN HARRIS PAID for by BIDEN FOR PRESIDENT
358329452153744	Joe Biden	0.15	No	Pro-Dem	We are the United States of America. There is not a single thing we cannot do. Are you with us? Join our campaign to elect Joe Biden today! BIDEN FOR PRESIDENT In 1994, Joe Biden secured the enactment of the Violence Against Women Act, which he authored. In the White House, he'll make sure it's reauthorized—and strengthened. VOTE VOTE TODAY! President Trump needs your vote. VOTE TODAY! LET'S GO
3832750618811	Donald J. Trump	0.18	No	Pro-Rep	

Ad ID	Page Name	TOXICITY	Toxicity > 0.5	Political Affiliation	(Extracted) Text
291746668466715	Joe Biden	0.59	Yes	Pro-Dem	<p>Donate at Joe Biden.com/victory. Stop Trump's Lies [Donate]</p> <p>The FINAL presidential debate before the election is this week and we already know Trump is going to spend his time lying to the American people. He'll dodge any question about his plans to address rising COVID-19 cases, unemployment, and relief for American families. Why? Because he has no plan. Don't let him fool you with his lies – not during this debate or for the next 4 years. Put an end to the lies by chipping in whatever you can to defeat Trump and elect me, Kamala, and Democrats nationwide</p>
680315525903929	Tony Gonzales	0.78	Yes	Pro-Rep	<p>During her last failed bid for Congress, Gina Jones publicly called for a debate — but when her challenge was accepted, she made excuse after excuse. The liberal media won't admit it, but Gina Jones is a hypocrite — plain and simple. TX23 deserves a leader with integrity, not another crooked radical with a personal agenda. #TeamTony #TX23 #WhereIsGina WhereIsGina.com GINA JONES SAYS DEBATE DODGING IS "NOTHING SHORT OF SHAMEFUL" NO ONE LIKES A HYPOCRITE</p>
688399545087658	Donald J. Trump	0.80	Yes	Pro-Rep	<p>Donate NOW BLATANT CORRUPTION The Democrats are the biggest hypocrites in the world. Crazy Nancy Pelosi is getting decimated for having a beauty parlor open in San Francisco when all others are forced to be closed. And to make it worse, she was caught on camera WITHOUT a face covering. Ridiculous! President Trump is calling on YOU to FIGHT BACK! We're giving him a list of supporters who step up in the NEXT HOUR to DEFEND their President and their Country. Please contribute ANY AMOUNT in the NEXT HOUR to fight for your Country and get on the list we give President Trump! Nancy Pelosi is Two-Faced DONATE NOW PAID FOR BY TRUMP MAKE AMERICA GREAT AGAIN COMMITTEE.</p>

Ad ID	Page Name	TOXICITY	Toxicity > 0.5	Political Affiliation	(Extracted) Text
256141942312712	Massachusetts Republican Party	0.89	Yes	Pro-Rep	The Radical Democrats support killing babies born alive during an abortion. We need to defeat the ROE Act and the Radical Democrats. THE DEMOCRATIC PARTY SUPPORTS KILLING BABIES BORN ALIVE
1046959635747679	RootsAction	0.96	Yes	Pro-Dem	Like Bernie says - Donald Trump is an idiot. He doesn't believe in climate change. He doesn't believe in anything that progressives fight for. Let's #VoteTrumpOut and get a worthwhile opponent in the White House. "DONALD TRUMP BELIEVES CLIMATE CHANGE IS A HOAX. DONALD TRUMP IS AN IDIOT" -BERNIE SANDERS

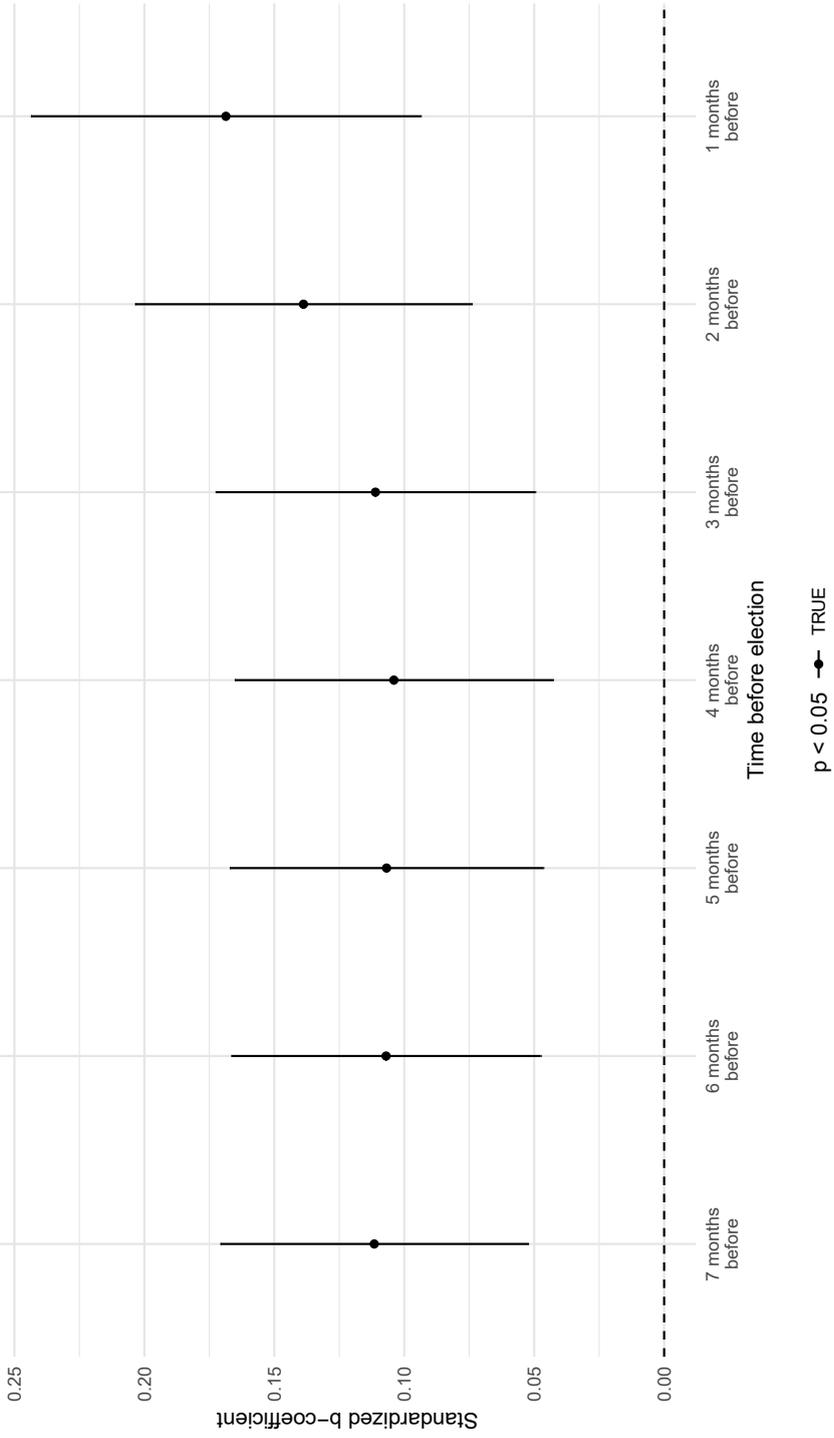
Figure A4
Specification Curve of Model 2 and 3; DV: Targeting (ordinal)





Simonsohn et al., 2015 propose visualizing results from the multiverse as a specification curve, which consists of two panels. The top panel shows the effect size (odds ratios for ordinal logistic regressions) for each universe (or specification503n). The bottom panel shows which specification of the parameters results in that result.

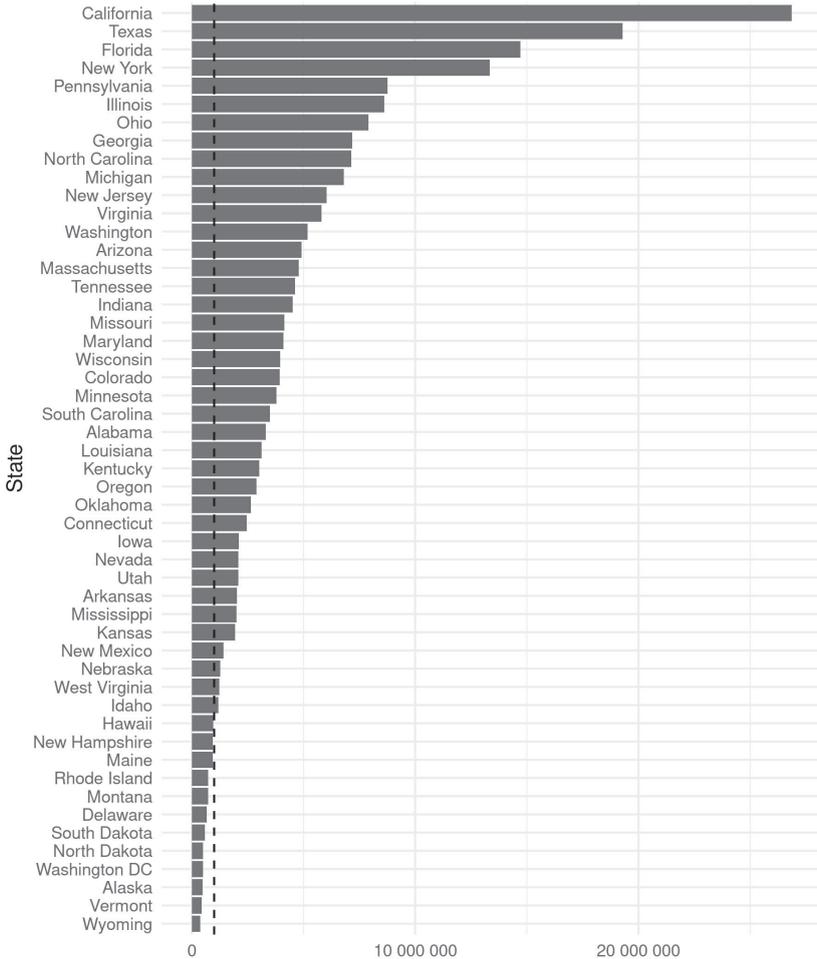
Figure A5
Specification Curve of Model 5; IV: Advertiser Type (Outside groups); DV Toxicity



The graph shows the effect size (standardized b-coefficients for linear regressions) for each time cut-off.

Figure A6

Estimated Facebook and Instagram users 13 years and older



Estimated Facebook and Instagram Users per State (13+)

Population estimates are from census.gov combined with ANES data for Facebook/Instagram use. Dashed line represents 1m estimated users.

Control variables

The following control variables are used in the various models presented here.

Political Affiliation (Pro-Democrat / Pro-Republican)

This variable is based on the meta-data matched from the FEC and CRP data. For the purposes of this investigation, we group the Biden campaign, Democrat politicians and (state) parties, pro-Democrat entities and committees as “Pro-Democrat”. Similarly, funding entities that support Trump for president (incl. the Trump campaign itself), Republican politicians and (local) parties as well as pro-Republican entities and committees as “Pro-Republican”.

Age Audience

While the Facebook Ad library does not provide info on who was targeted with an ad, they do provide *audience shares* for sub-regions (in the US case: states), age and gender. This variable represents the share of the audience that saw the ad in percent. If an ad was not seen by a particular audience at all, that audience group will not be listed for that ad. This way, we can create a categorical variable based on the audience groups that saw the ad. For *Age Audience*, we create 5 separate groups:

- All Ages (reference group): 13-65+
- Younger Audience: 13-34
- Middle-Aged Audience: 35-54
- Older Audience: 55+
- Other: All other combinations (e.g. younger and older audiences together)

Gender Audience

This is similar to *Age Audience*, only that it is a categorical variable showing whether an ad was seen by men, women or both.

- Female (reference group)
- Male & Female
- Male

Unique Ad runs

This variable measures the number of times that an *unique* ad ran. An ad is considered unique when it shares the same text (which includes text

extracted from images and videos) and was run by the same advertiser. The idea is to control for the fact that ads run more often given that the same ad may run at different points of time. Take for example these two ads by the Facebook page “Donald J. Trump”

1. https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US&q=2671672779766119&sort_data%5bdirection%5d=desc&sort_data%5bmode%5d=relevancy_monthly_grouped&search_type=keyword_unordered&media_type=all
2. [https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US&q=592114724810931&sort_data\[direction\]=desc&sort_data\[mode\]=relevancy_monthly_grouped&search_type=keyword_unordered&media_type=all](https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US&q=592114724810931&sort_data[direction]=desc&sort_data[mode]=relevancy_monthly_grouped&search_type=keyword_unordered&media_type=all)

Both ads are identical and run by the same page, but ran twice several days apart. We count each time this unique ad ran as one “unique ad run”.

Ad run time

This variable is similar but it measures the number of days that an unique ad ran. So if an ad ran from August 1st to August 3rd 2020 and then again from September 1st to September 3rd 2020 it ran for a total of 6 days.

Closeness to Election

This variable measures closeness to the election (in days). The greater the number, the closer is election day (November 3rd 2020).

Spending

How much money was spent on an ad. As with potential reach, Facebook provides the spending measure in lower and upper bounds. However, it is interesting to note that the Facebook Ad Library API documentation differs from what was actually found in the data: the documentation states that the measures comes in 10 ranges of dollars spent “<100, 100-499, 500-999, 1K-5K, 5K-10K, 10K- 50K, 50K-100K, 100K-200K, 200K-500K, >1M”, which is contradicted by the data that we were able to retrieve from the API, which shows that the values of spending are in fact more granular, even if still within a discrete set of values (45 distinct ranges as opposed to the 10 ranges specified in the documentation). The measure here represents the lower boundary because the upper boundary has an open category. For the lowest

lower boundary (\$0) we set the minimum amount of dollars spent per ad to \$1 since we can reasonable assume placing ads always infers a minimum cost.

Maximum users reached

Maximum number of reached Facebook and Instagram users inferred from the latest round of the American National Election Survey (ANES, 2021) as well as 2019 age and gender population estimates per state taken from the US Census Bureau (US Census Bureau, 2020). Using ANES, we calculate the percentage of Facebook users for each age and gender group.⁴

With the help of this metric we can then calculate the actual number of users using population estimates by the US Census Bureau whereby we assume that age and gender Facebook uptake does not significantly differ between US states.⁵ Next, we make use of the aforementioned *audience share* variables: for each ad we have a string of age and gender groups that saw the ad in addition to in which state(s) the ad was shown. We combine this information with the Facebook user estimates by age and gender and per US state. So for example, if an ad was only seen by women of all age groups in Alabama and North Carolina, then we would add up the number of all female Facebook users of all age groups in Alabama and North Carolina and conclude that this is the maximum number of users that the ad *could have reached*.

We control for maximum users reached because in our data we have campaigns that may target different geographical regions systematically: for example, a page may only target a specific state and given that our main DV for H1 (potential reach) has fixed levels of potentially reached people, the comparison of Page A targeting 500,000 people in every state of the country vs. Page B targeting 500,000 people because they target a specific state or region might be flawed. To account for that, we set the potential reach measure in relation to the maximum number of people the ad could have reached by using it as a control variable.

Codebook Political Incivility

This codebook is adapted from (Vargo & Hopp, 2020) who used this for manually annotating ads for incivility from the Russian-linked disinformation campaign.

Code attributes as “1” if present, “0” if absent. Follow the following definitions:

Identity-Based Negative Language. This category deals with attacks that either uses identity features (e.g., race, sexual orientation, gender,

immigration status) in a negative manner or attempts to negatively situate two or more identity groups. Positive mentions of identity are not considered identity-based negative (e.g., celebrating Gay Pride).

Inflammatory Language. This category speaks to attack that use of unnecessarily negative emotional language that urges audiences to be upset and/or to take action regarding a political or social issue (in reference to a political opponent).

Obscene Language. If an attack contains any vulgar language, including any known “swear” words, no matter how mild, including “hell” or “damn,” this attribute is present.

Even quoted language should be considered. Identity labels used in positive ways (e.g., gay, or Muslim) are not obscene. Attempts to censor swear words (e.g., bullsh*t) are still obscenity.

Threatening Language. If a potential threat is exposed to the audience, then this attribute is present. For instance, warnings of police violence, warnings of death because of mismanagement of COVID-19 pandemic, or warnings of the risks associated with illegal immigration are all examples of threats to the audience. Frequently, these threats will be suggested or implied, not directly posed to the audience.

If any of the four sub forms of incivility is present we consider an ad to be uncivil.

Validity checks

It is worth checking how the Perspective API performs in detecting incivility in our sample of ads. Table A1 in the appendix lists ten randomly chosen ads and reveals decent face validity of the measure: ads that score higher in toxicity (i.e. are closer to 1) include accusations of political opponents being liars, hypocrites, idiots, corrupt and even accuses them of “baby-killing” when referring to abortion. Further, Figure A2 reveals that toxicity is associated with negative sentiment, and Figure A3 shows that toxic ads are more likely to include mentions of opponents compared to non-toxic ads (both figures are found in appendix). We also performed manual coding in which two researchers coded a sample of 120 ads for *civil* and *uncivil* ads (see codebook in appendix). Pairwise agreement between coders and binary toxicity predicted from the Perspective API was 82.5% and 81.7% for coder 1 and 2, respectively. An intercoder-reliability of Krippendorff’s Alpha between all three codings shows a value of 0.704, above the minimum acceptable limit of 0.66 but below the optimal standard of above 0.8. Nonetheless, for our intents we can conclude that the Perspective API gives acceptable measures of incivility in political advertisements.

Correlations & Multicollinearity checks

Table A2

VIF values for Model 2

Coefficient	VIF	SE Factor	Tolerance
Toxicity	1.01	1.00	0.99
Political Affiliation	1.02	1.01	0.98
Advertiser Type: Outside groups	1.02	1.01	0.98
log. Closeness to Election	1.10	1.05	0.91
log. Spending (lower bound)	1.10	1.05	0.91
log. Ad runs	1.31	1.14	0.76
log. Ad run time	1.36	1.17	0.74
log. Max Reached	1.52	1.23	0.66
Gender	1.57	1.25	0.64
Age Group	1.86	1.36	0.54

VIF (variance inflation factor) values for all variables are well below problematic ranges of > 5.

VIF values for all variables are well below problematic ranges of > 5.

Figure A7
Correlation between study variables

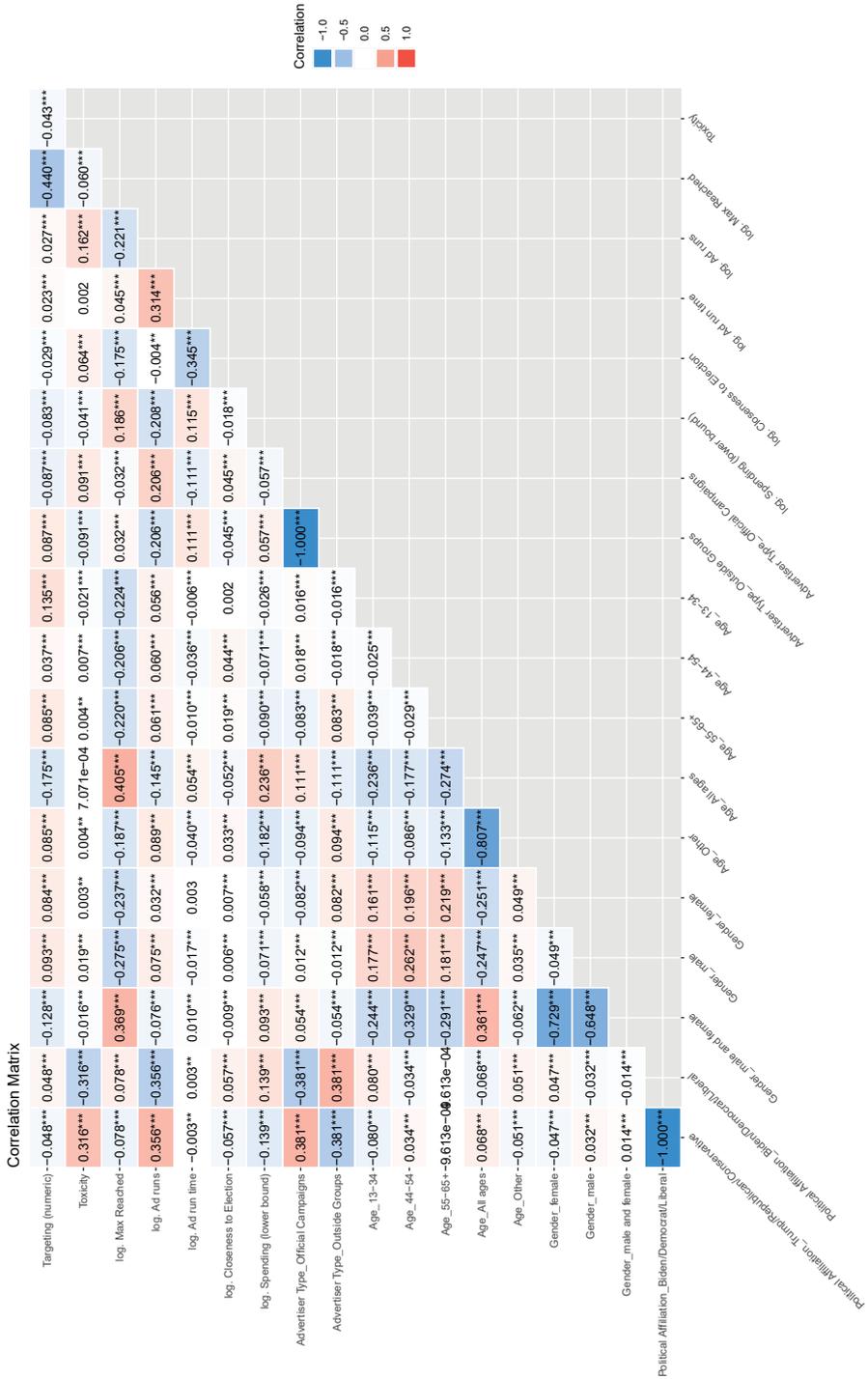


Table A3*VIF values for Model 5*

Coefficient	VIF	SE Factor	Tolerance
Political Affiliation	1.02	1.01	0.98
Advertiser Type: Outside groups	1.02	1.01	0.98
log. Closeness to Election	1.09	1.04	0.92
log. Spending (lower bound)	1.10	1.05	0.91
log. Ad runs	1.31	1.14	0.77
log. Ad run time	1.36	1.16	0.74
log. Max Reached	1.52	1.23	0.66
Gender	1.57	1.25	0.64
Age Group	1.86	1.36	0.54

References

- Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–30. <https://doi.org/10.1145/3359301>
- Ali, M., Sapiezynski, P., Korolova, A., Mislove, A., & Rieke, A. (2021). Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 13–21. <https://doi.org/10.1145/3437963.3441801>
- ANES. (2021). 2020 Time Series Study. Retrieved April 26, 2021, from <https://electionstudies.org/data-center/2020-time-series-study/>
- Ansolabehere, S., Iyengar, S., Simon, A., & Valentino, N. (1994). Does Attack Advertising Demobilize the Electorate? [Publisher: [American Political Science Association, Cambridge University Press]]. *The American Political Science Review*, 88(4), 829–838. <https://doi.org/10.2307/2082710>
- Anstead, N. (2017). Data-Driven Campaigning in the 2015 United Kingdom General Election. *The International Journal of Press/Politics*, 22(3), 294–313. <https://doi.org/10.1177/1940161217706163>
- Barbu, O. (2014). Advertising, Microtargeting and Social Media. *Procedia – Social and Behavioral Sciences*, 163, 44–49. <https://doi.org/10.1016/j.sbspro.2014.12.284>
- Baum, L. (2020). Presidential General Election Ad Spending Tops \$1.5 Billion | Wesleyan Media Project. Retrieved March 28, 2021, from <https://mediaproject.wesleyan.edu/releases-102920/>

- Budge, I., & Farlie, D. (1977). *Voting and party competition: A theoretical critique and synthesis applied to surveys from ten democracies*. Wiley.
- Chand, D. E. (2017). Dark Money” and “Dirty Politics”: Are anonymous ads more negative? *Business and Politics*, 19(3), 454–481. <https://doi.org/10.1017/bap.2016.13>
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, 64(4), 658–679. <https://doi.org/10.1111/jcom.12104>
- Confessore, N. (2018). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*. Retrieved April 13, 2021, from <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>
- Constine, J. (2019). Facebook launches searchable transparency library of all active ads. Retrieved April 30, 2021, from <https://social.techcrunch.com/2019/03/28/facebook-ads-library/>
- Coordination Watch. (2022). Coordination Watch. Retrieved April 7, 2022, from <https://www.coordinationwatch.org>
- CRP. (2021a). 2020 Outside Spending, by Group | OpenSecrets. Retrieved April 25, 2021, from <https://www.opensecrets.org/outsidespending/summ.php?cycle=2020&disp=O&type=A&chrt=D>
- CRP. (2021b). Online Political Ad Spending. Retrieved April 25, 2021, from <https://www.opensecrets.org/online-ads>
- CRP. (2022). Outside Spending | OpenSecrets. Retrieved April 19, 2022, from <https://www.opensecrets.org/outsidespending/>
- Damore, D. F. (2002). Candidate Strategy and the Decision to Go Negative [Publisher: SAGE Publications Inc]. *Political Research Quarterly*, 55(3), 669–685. <https://doi.org/10.1177/106591290205500309>
- Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2020). Do (Micro-targeted) Deepfakes Have Real Effects on Political Attitudes? [Publisher: SAGE Publications Inc]. *The International Journal of Press/Politics*, 1940161220944364. <https://doi.org/10.1177/1940161220944364>
- Dobber, T., Ó Fathaigh, R., & Zuiderveen Borgesius, F. J. (2019). The regulation of online political micro-targeting in Europe. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1440>
- Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. H. (2017). Two crates of beer and 40 pizzas: The adoption of innovative political behavioural targeting techniques. *Internet Policy Review*, 6(4). <https://doi.org/10.14763/2017.4.777>
- Dommett, K. (2019). Data-driven political campaigns in practice: Understanding and regulating diverse data-driven campaigns. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1432>

- Downs, A. (1957). *An economic theory of democracy*. Harper Collins. Edyvane, D. (2020). Incivility as Dissent. *Political Studies*, 68(1), 93–109. <https://doi.org/10.1177/0032321719831983>
- Facebook Business Help Center. (2021). About Potential Reach. Retrieved May 5, 2021, from <https://www.facebook.com/business/help/1665333080167380>
- FEC. (2021). Committees. Retrieved April 25, 2021, from <https://www.fec.gov/data/committees/?cycle=2020>
- Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, 38(6), 889–906. <https://doi.org/10.1037/0022-3514.38.6.889>
- Fowler, E. F., Franz, M. M., Martin, G. J., Peskowitz, Z., & Ridout, T. N. (2021). Political Advertising Online and Offline. *American Political Science Review*, 115(1), 130–149. <https://doi.org/10.1017/S0003055420000696>
- Frederic, N. S., & Falomir-Pichastor, J. M. (2018). Heterogeneity of Ingroup Identity and Anti-Immigrant Prejudice: The Moderating Role of RWA and Outgroup Homogeneity. *International Review of Social Psychology*, 31(1). <https://doi.org/10.5334/irsp.152>
- Frimer, J. A., & Skitka, L. J. (2018). The Montagu Principle: Incivility decreases politicians' public approval, even with their political base. *Journal of Personality and Social Psychology*, 115(5), 845–866. <https://doi.org/10.1037/pspi000140>
- Garramone, G. M. (1984). Voter Responses to Negative Political Ads. *Journalism Quarterly*, 61(2), 250–259. <https://doi.org/10.1177/107769908406100202>
- Geer, J. G. (2006). *In defense of negativity: Attack ads in presidential campaigns* [OCLC: ocm60589072]. University of Chicago Press.
- Gervais, B. T. (2019). Rousing the Partisan Combatant: Elite Incivility, Anger, and Antideliberative Attitudes. *Political Psychology*, 40(3), 637–655. <https://doi.org/10.1111/pops.12532>
- Ghosh, A., Venkatadri, G., & Mislove, A. (2019). Analyzing political advertisers' use of facebook's targeting features. *IEEE workshop on technology and consumer protection (ConPro'19)*.
- Google Cloud Vision API. (2022). Vision AI | Derive Image Insights via ML | Cloud Vision API. Retrieved May 22, 2022, from <https://cloud.google.com/vision>
- Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., & Ng, A. Y. (2014). Deep Speech: Scaling up end-to-end speech recognition [arXiv:1412.5567]. *arXiv:1412.5567 [cs]*. Retrieved January 30, 2021, from <http://arxiv.org/abs/1412.5567>
- Haselmayer, M. (2019). Negative campaigning and its consequences: A review and a look ahead. *French Politics*, 17 (3), 355–372. <https://doi.org/10.1057/s41253-019-00084-8>
- Herbst, S. (2010). *Rude democracy: Civility and incivility in American politics* [OCLC:ocn557406946]. Temple University Press.

- Hersh, E. D., & Schaffner, B. F. (2013). Targeted Campaign Appeals and the Value of Ambiguity. *The Journal of Politics*, 75(2), 520–534. <https://doi.org/10.1017/S0022381613000182>
- Hilbig, B. E. (2009). Sad, thus true: Negativity bias in judgments of truth. *Journal of Experimental Social Psychology*, 45(4), 983–986. <https://doi.org/10.1016/j.jesp.2009.04.012>
- Hill, R. P. (1989). An Exploration of Voter Responses to Political Advertisements. *Journal of Advertising*, 18(4), 14–22. <https://doi.org/10.1080/00913367.1989.10673163>
- Hopp, T., Vargo, C. J., Dixon, L., & Thain, N. (2020). Correlating Self-Report and Trace Data Measures of Incivility: A Proof of Concept. *Social Science Computer Review*, 38(5), 584–599. <https://doi.org/10.1177/0894439318814241>
- Kim, Y. M., Hsu, J., Neiman, D., Kou, C., Bankston, L., Kim, S. Y., Heinrich, R., Baragwanath, R., & Raskutti, G. (2018). The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication*, 35(4), 515–541. <https://doi.org/10.1080/10584609.2018.1476425>
- Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.
- Krotzek, L. J. (2019). Inside the Voter's Mind: The Effect of Psychometric Microtargeting on Feelings Toward and Propensity to Vote for a Candidate. *International Journal of Communication; Vol 13 (2019)*. <https://ijoc.org/index.php/ijoc/article/view/9605>
- Krupnikov, Y. (2011). When Does Negativity Demobilize? Tracing the Conditional Effect of Negative Campaigning on Voter Turnout [Publisher: [Midwest Political Science Association, Wiley]]. *American Journal of Political Science*, 55(4), 797–813. Retrieved May 4, 2021, from <http://www.jstor.org/stable/23025121>
- Krupnikov, Y. (2015). Kyle Mattes and David P. Redlawsk. *The Positive Case for Negative Campaigning*. Chicago: University of Chicago Press. 2014. 256 pp. \$75.00 (cloth). \$25.00 (paper). *Public Opinion Quarterly*, 79(3), 825–827. <https://doi.org/10.1093/poq/nfv032>
- Krupnikov, Y., & Piston, S. (2015). Accentuating the Negative: Candidate Race and Campaign Strategy. *Political Communication*, 32(1), 152–173. <https://doi.org/10.1080/10584609.2014.914612>
- Kruschinski, S., & Haller, A. (2017). Restrictions on data-driven political micro-targeting in Germany. *Internet Policy Review*, 6(4). <https://doi.org/10.14763/2017.4.780>
- Lau, R. R., & Rovner, I. B. (2009). Negative Campaigning. *Annual Review of Political Science*, 12(1), 285–306. <https://doi.org/10.1146/annurev.polisci.10.071905.101448>
- Lau, R. R., Sigelman, L., & Rovner, I. B. (2007). The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment. *The Journal of Politics*, 69(4), 1176–1209. <https://doi.org/10.1111/j.1468-2508.2007.00618.x>

- Lavigne, M. (2020). Strengthening ties: The influence of microtargeting on partisan attitudes and the vote. *Party Politics*, 135406882091838. <https://doi.org/10.1177/1354068820918387>
- López Ortega, A. (2021). Are microtargeted campaign messages more negative and diverse? An analysis of Facebook Ads in European election campaigns. *European Political Science*. <https://doi.org/10.1057/s41304-021-00346-6>
- Los Angeles Times. (2012). Attack ads to blast Mitt Romney's work at Bain Capital. Retrieved April 30, 2022, from <https://www.latimes.com/la-xpm-2012-jan-07-lana-0108-gingrich-romney-bain-20120108-story.html>
- Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The Effects of Negativity and Motivated Information Processing During a Political Campaign. *Journal of Communication*, 56(1), 27–51. <https://doi.org/10.1111/j.1460-2466.2006.00003.x>
- Metcalfe, A. L., Angle, J. W., Phelan, C. N., Muth, B. A., & Finley, J. C. (2018). More “Bank” for the Buck: Microtargeting and Normative Appeals to Increase Social Marketing Efficiency [Publisher: SAGE Publications Inc]. *Social Marketing Quarterly*, 25(1), 26–39. <https://doi.org/10.1177/1524500418818063>
- Muddiman, A. (2017). Personal and Public Levels of Political Incivility [Number: 0]. *International Journal of Communication*, 11(0), 21. Retrieved May 2, 2022, from <https://ijoc.org/index.php/ijoc/article/view/6137>
- Mutz, D. C., & Reeves, B. (2005). The New Videomalaise: Effects of Televised Incivility on Political Trust. *American Political Science Review*, 99(1), 1–15. <https://doi.org/10.1017/S0003055405051452>
- Nai, A., & Maier, J. (2020). Is Negative Campaigning a Matter of Taste? Political Attacks, Incivility, and the Moderating Role of Individual Differences. *American Politics Research*, 1532673X2096554. <https://doi.org/10.1177/1532673X20965548>
- Pascual-Ferrá, P., Alperstein, N., Barnett, D. J., & Rimal, R. N. (2021). Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic [Publisher: SAGE Publications Ltd]. *Big Data & Society*, 8(1), 20539517211023533. <https://doi.org/10.1177/20539517211023533>
- Perspective API. (2021). About the API – Attributes and Languages. Retrieved May 5, 2021, from <https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages>
- Pew Research Center. (2020). Voters' reflections on the campaign. Retrieved May 2, 2022, from <https://www.pewresearch.org/politics/2020/11/20/voters-reflections-on-the-campaign/>
- Raynauld, V., & Turcotte, A. (2018). Different Strokes for Different Folks”: Implications of Voter Micro-Targeting and Appeal in the Age of Donald Trump. In J. Gillies (Ed.), *Political Marketing in the 2016 U.S. Presidential Election* (pp. 11–28). Springer International Publishing. https://doi.org/10.1007/978-3-319-59345-6_2

- Ridout, T. N., Franz, M. M., & Fowler, E. F. (2015). Sponsorship, Disclosure, and Donors: Limiting the Impact of Outside Group Ads [Publisher: SAGE Publications Inc]. *Political Research Quarterly*, 68(1), 154–166. <https://doi.org/10.1177/1065912914563545>
- Riker, W. H. (1991). Why Negative Campaigning Is Rational: The Rhetoric of the Ratification Campaign of 1787–1788. *Studies in American Political Development*, 5(2), 224–283. <https://doi.org/10.1017/S0898588X00000249>
- Riker, W. H., Calvert, R. L., Mueller, J. E., & Wilson, R. K. (1996). *The strategy of rhetoric: Campaigning for the American Constitution*. Yale University Press.
- Roberts, C. (2013). A Functional Analysis Comparison of Web-Only Advertisements and Traditional Television Advertisements from the 2004 and 2008 Presidential Campaigns. *Journalism & Mass Communication Quarterly*, 90(1), 23–38. <https://doi.org/10.1177/1077699012468741>
- Robertson, D. B. (1976). *A theory of party competition*. J. Wiley.
- Rohrschneider, R. (2002). Mobilizing versus chasing: How do parties target voters in election campaigns? *Electoral Studies*, 21(3), 367–382. [https://doi.org/10.1016/S0261-3794\(00\)00044-5](https://doi.org/10.1016/S0261-3794(00)00044-5)
- Rossini, P. (2020). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, 009365022092131. <https://doi.org/10.1177/0093650220921314>
- Rossini, P. (2021). More Than Just Shouting? Distinguishing Interpersonal-Directed and Elite-Directed Incivility in Online Political Talk. *Social Media + Society*, 7(2), 205630512110088. <https://doi.org/10.1177/20563051211008827>
- Roy, J., & Alcantara, C. (2016). Fighting Fire with Fire: The Implications of (Not) Going Negative in a Multiparty Election Campaign. *Canadian Journal of Political Science*, 49(3), 473–497. <https://doi.org/10.1017/S0008423916000548>
- Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, 5(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2
- Scala, D. J. (2014). Are Super PACs Arms of Political Parties? A Study of Coordination. In R. W. Holder & P. B. Josephson (Eds.), *The American Election 2012: Contexts and Consequences* (pp. 69–80). Palgrave Macmillan US. https://doi.org/10.1057/9781137389220_7
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2694998>
- Sobieraj, S., & Berry, J. M. (2011). From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News. *Political Communication*, 28(1), 19–41. <https://doi.org/10.1080/10584609.2010.542360>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>

- Stuckelberger, S. (2019). Mobilizing and chasing: The voter targeting of negative campaigning – Lessons from the Swiss case. *Party Politics*, 135406881985225. <https://doi.org/10.1177/1354068819852252>
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1410>
- Sydnor, E. (2019). *Disrespectful democracy: The psychology of political incivility*. Columbia University Press.
- Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The Dynamics of Political Incivility on Twitter. *SAGE Open*, 10(2), 215824402091944. <https://doi.org/10.1177/2158244020919447>
- Theocharis, Y., & Jungherr, A. (2021). Computational Social Science and the Study of Political Communication [Publisher: Routledge]. *Political Communication*, 38(1-2), 1–22. <https://doi.org/10.1080/10584609.2020.1833121>
- US Census Bureau. (2020). 2019 Population Estimates by Age, Sex, Race and Hispanic Origin [Section: Government]. Retrieved April 26, 2021, from <https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html>
- Vaithyanathan, D., & Muniraj, M. (2019). Cloud based Text extraction using Google Cloud Vision for Visually Impaired applications. *2019 nth International Conference on Advanced Computing (ICoAC)*, 90–96. <https://doi.org/10.1109/ICoAC48765.2019.246822>
- Valkenburg, P. M., & Peter, J. (2013). The Differential Susceptibility to Media Effects Model: Differential Susceptibility to Media Effects Model. *Journal of Communication*, 63(2), 221–243. <https://doi.org/10.1111/jcom.12024>
- Valli, C., & Nai, A. (2020). Attack politics from Albania to Zimbabwe: A large-scale comparative study on the drivers of negative campaigning. *International Political Science Review*, 019251212094641. <https://doi.org/10.1177/0192512120946410>
- van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science [Publisher: Routledge]. *Communication Methods and Measures*, 12(2-3), 81–92. <https://doi.org/10.1080/19312458.2018.1458084>
- Vargo, C. J., & Hopp, T. (2020). Fear, Anger, and Political Advertisement Engagement: A Computational Case Study of Russian-Linked Facebook and Instagram Content. *Journalism & Mass Communication Quarterly*, 97 (3), 743–761. <https://doi.org/10.1177/1077699020911884>
- Votta, F. (2021). Favstats/peRpective [original-date: 2019-05-06T21:53:53Z]. Retrieved May 6, 2021, from <https://github.com/favstats/peRpective>
- Wagner, K. (2017). Donald Trump and Hillary Clinton spent \$81 million on Facebook ads before last year's election. Retrieved March 28, 2021, from <https://www.vox.com/2017/11/1/16593066/trump-clinton-facebook-advertising-money-election-president-russia>

- Wakabayashi, D. (2017). Google Cousin Develops Technology to Flag Toxic Online Comments. *The New York Times*. Retrieved April 5, 2021, from <https://www.nytimes.com/2017/02/23/technology/google-jigsaw-monitor-toxic-online-comments.html>
- Walter, A. S., & van der Eijk, C. (2019). Unintended consequences of negative campaigning: Backlash and second-preference boost effects in a multi-party context. *The British Journal of Politics and International Relations*, 21(3), 612–629. <https://doi.org/10.1177/1369148119842038>
- Who Targets Me. (2020). How to get to “gold standard” political ad transparency. Retrieved May 5, 2021, from <https://whotargets.me/how-to-take-a-gold-standard-approach-to-political-advertising-transparency-and-policy/>
- Who Targets Me. (2021). WhoTargetsMe/ad-transparency-standards [original-date: 2020-12-16T11:40:34Z]. Retrieved May 5, 2021, from <https://github.com/WhoTargetsMe/ad-transparency-standards>
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01832>
- Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*, 009365022096196. <https://doi.org/10.1177/0093650220961965>
- Zuiderveen Borgesius, F., Moeller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodó, B., & de Vreese, C. H. (2018). *Online Political Microtargeting: Promises and Threats for Democracy* (SSRN Scholarly Paper No. ID 3128787). Social Science Research Network. Rochester, NY. Retrieved September 10, 2020, from <https://papers.ssrn.com/abstract=3128787>