Manipulative tactics are the norm in political emails
Evidence from 100K emails from the 2020 U.S. election cycle

Arunesh Mathur
Princeton University
amathur@cs.princeton.edu

Angelina Wang
Princeton University
angelina.wang@princeton.edu

Carsten Schwemmer
GESIS - Leibniz-Institute for the Social Sciences
c.schwem2er@gmail.com

Maia Hamin
Princeton University
mhamin@princeton.edu

Brandon M. Stewart
Princeton University
bms4@princeton.edu

Arvind Narayanan
Princeton University
arvindn@cs.princeton.edu

For the most recent version of this working paper, please visit https://electionemails2020.org

Paper date: October 5, 2020

Abstract
Manipulative political discourse undermines voters’ autonomy and thus threatens democracy. Using a newly assembled corpus of more than 100,000 political emails from over 2,800 political campaigns and organizations sent during the 2020 U.S. election cycle, we find that manipulative tactics are the norm, not the exception. The majority of emails nudge recipients to open them by employing at least one of six manipulative tactics that we identified; the median sender uses such tactics 43% of the time. Some of these tactics are well known, such as sensationalistic subject lines. Others are more devious, such as deceptively formatted “From:” lines that attempt to trick recipients into believing that the message is a continuation of an ongoing conversation. Manipulative fundraising tactics are also rife in the bodies of emails. Our data can be browsed at electionemails2020.org

Introduction
Political campaigns use a variety of digital media for mobilizing potential voters and raising funds. Depending on the medium, campaigns adjust the style and substance of their messages, take advantage of technological possibilities, and adapt to limitations [1] [2] [3]. In that regard, email is a particularly important medium: it is a major driver of grassroots fundraising and requires little campaign infrastructure to operate [2]. Further, emails can be tailored rapidly in response to news [2] and campaign staffers rate them as more representative of campaign strategy than either television advertisements or media coverage [4]. Yet, due to their semi-public
nature, campaign emails are difficult to observe at scale and have consequently received relatively little study.

One of these studies analyzes a small set of campaign emails from the 2004 U.S. presidential election, finding that political emails have used tactics common in the business world, such as viral marketing. In the 2020 cycle, there has been a spate of journalistic and anecdotal reports of manipulative tactics such as misleading subject lines and deceptive fundraising techniques including nonexistent deadlines and false claims of donation matching. However, these reports are based on small samples of emails and senders. The prevalence and variety of manipulative political email tactics aren’t yet well understood. These questions are vital to study because manipulative political discourse undermines voters’ autonomy, generates cynicism and thus threatens democracy. Following the Cambridge Analytica scandal, the potential for political manipulation through micro-targeted online advertising has been recognized, but email has largely escaped scrutiny.

Assembling a corpus of 100,000 political emails

Our corpus, which will continue updating through the end of the 2020 campaign, currently contains more than 250K emails from more than 3000 political campaigns and organizations in the 2020 election cycle in the United States. It includes candidates in federal and state races as well as Political Action Committees (PACs), Super PACs, political parties, and other political organizations. Data collection began in December 2019. Our corpus has two orders of magnitude more emails than the largest corpus of election-related emails previously analyzed in the academic literature.

The relevant population is constantly changing as candidates announce and end their candidacy and organizations enter the fray at different points. To maximize completeness, we combined three sources of information about political entities. We purchased a list from Ballotpedia — updated every week — of candidates running for election in 2020 at the federal and state levels. We gathered a list of active PACs and other groups from OpenSecrets. Finally, we compiled a list of political and Hill committees for political parties at the federal and state level based on the fact that their names conform to a standard pattern (e.g., “Young Democrats/Republicans” of America”).

Then, we associated each entity with its website, if one existed. If Ballotpedia recorded a website, or if one was recorded in the FEC filings, we used that information. Otherwise we queried search engines Bing and DuckDuckGo based on the entity’s name and office and ranked the results based on a number of heuristics.
that we developed by observing common patterns in the URLs and names of such websites (e.g., “<last name> for Congress”). We manually verified each website that was detected with our automatic procedure.

Next, we automated the process of finding and filling out email subscription forms on these websites. We created a bot based on Englehardt et. al.’s open-source code [16, 17]. For each website, if the bot discovers an email sign-up form, it fills it in with the information of a fictional recipient. The bot generates a different, unique email address for each form, and monitors all the addresses for incoming email. Since our list of campaigns and organizations grew over time, we executed the website discovery and automated subscription steps in seven waves starting in December 2019.

On receiving an email, the bot opens it exactly once and clicks on the confirmation link, if one is present. It also downloads all remote resources—including cookies and tracking pixels—embedded in the email and takes a screenshot of the email. These form the contents of our corpus.

Overview of the corpus

In this paper, we analyze the 100K emails sent by over 2,800 senders that were collected from December 2019 to June 25, 2020. Figure 1 shows the daily volume of emails. It exhibits weekly cycles, monthly spikes, as well as spikes resulting from welcome emails following waves of online subscription. The gradual increase in daily volume is due both to the waves of subscription and due to the increase in volume per sender as the election cycle progresses. The 2,834 senders include about 1,084 federal candidates, 1,359 state candidates, 264 PACs (including super PACs and hybrid PACs), and 127 other organizations. Senders differ greatly in how often they send emails: the median campaign sends only an email every three weeks, whereas the 90th percentile campaign sends three emails per week. PACs and organizations are more active: the median sends an email every two weeks and the 90th percentile sends 6 emails per week.

The discrepancy between the number of entities that we started with (16,062) and the number of senders in our corpus (2,834) primarily reflects the fact that many entities don’t have active email campaigns, rather than a limitation of our automated signup. We give a few lines of evidence for this claim. First, of a sample of 91 websites on which we were able to successfully sign up manually, the bot succeeded on 80. The majority of failures (8 out of 11) were due to the bot’s confusion between form fields such as ZIP code and phone number, as these forms are not designed to be machine readable.

Second, our coverage of prominent candidates is substantially higher than the baseline: of the 23 candidates who participated in the democratic presidential debates, our corpus includes all 20 candidates who were still in the race when we started our data collection; similarly, it includes 86.4% of federal candidates and 70.6% organizations who raised more than 10 million dollars according to FEC filings. Of the organizations who did not send us emails, 51.9% raised no money, whereas among those who did, this fraction is much lower at 27.6%. Our coverage includes a diversity of electoral races: 89% of federal races and 45% of state races have at least one candidate who has sent us emails.

Third, we compared our corpus to politicalemails.org which is, to our knowledge, the only other large-scale political email collection effort for the 2020 U.S. electoral cycle. Of the sender types that are in scope for both efforts (i.e., excluding Non-Governmental Organizations, 501 organizations, and international organizations that are in scope for politicalemails.org), our corpus misses only 1 senders out of a sample of 30 senders in their corpus; conversely, only 10.6% of the senders in our corpus are present in theirs.

We note a few limitations of our coverage. As mentioned above, we don’t cover all entities that shape the political debate. Further, we only observe emails that result from signing up on campaign websites. It is possible that emails sent to lists acquired through other means—data brokers, in-person fundraisers, rallies—have substantially different content. Donors may also receive tailored content, which we do not
Figure 2: Subfigure A: overview of assigned topic categories for political emails. The majority of content is related to political campaigns, political actors and political events as well as political issues. More than 18% of content is related to explicit fundraising. Subfigures B-D: selected topics and their prevalence over time by party association. Emails of senders with other or unknown party affiliations are not included. Prevalence estimates are based on a line regression with spline basis functions. Error bands show 95% uncertainty intervals which include a global approximation to the uncertainty. Because modeling is done in an unconstrained space, the edges of the intervals can fall slightly below 0%.
observe. Geographic targeting and A/B testing may also result in variations of email content. For 24 prominent presidential candidates, we signed up with 50 distinct and unique email addresses, each time submitting a ZIP code from a different U.S. state, enabling us to detect geographic targeting as well as A/B testing. We leave this analysis for future iterations of our corpus.

The findings we present in this paper are based on a snapshot of our corpus as of June 25, 2020, including a total of 106,342 political emails. However, we have already accumulated and released over 250K emails and will continue collecting content through the end of the 2020 U.S. elections.

To provide an overview of the content of the emails we use a structural topic model [18]. Topic models discover latent themes (topics) from text documents. Each email is a proportional mix of topics, e.g. an email might be 70% “fundraising” and 30% about “volunteering,” the former consisting of words such as “donation” and “deadline” and the latter consisting of words such as “outreach” and “help”. Structural topic models further allow analyzing how these proportions vary in expectation depending on contextual information — in our case, party, type of sender (campaign vs. organization), date, and an interaction term between party and date.

To prepare the emails for topic modelling, we took two steps to prevent sender-specific terms from dominating the output: we stripped all non-textual content such as HTML tags, removed greetings, footers, and other text repeated across emails by the same sender; we also masked the name of the sender (candidate or organization). Additionally, we took standard steps such as lemmatization, identification of collocations terms, and removal of stopwords. We selected a model with 65 topics that best fits our research goals (more details available in our supplemental material) and further binned them into six high-level categories (Figure 2.A).

Topics in the “political campaigns, actors and events” category include, for example, President Trump, Joe Biden and control of the U.S. Senate. Prominent political issues include healthcare, the Black Lives Matter movement, LGBTQ-related issues and the Coronavirus pandemic. The third biggest category is explicit fundraising and donation requests (about 18%). However, this is a conservative estimate, as many of the email footers we removed during preprocessing also included fundraising information. Figures 2.B-D explore three topics that have been prominent in the national conversation during the observation period are seen as politically polarized. For COVID-19, the prevalence of the topic increased in March 2020 and reached a peak shortly after the U.S. national emergency declaration. During this time, about 8% of email content in our corpus were related to this topic. Differences by party are relatively minor for COVID-19 whereas the other two topics show more substantial differences along the expected lines. Additional figures are included in the supplementary material, showing similar differences along party lines for topics such as guns, education and foreign policy.

**Approaches to fundraising fatigue**

Fundraising is one of the main goals of campaign emails [1]. In addition to explicit fundraising, we found that the majority of emails (61 in a sample of 100) contain donation requests, even after we exclude footers which often contain such requests as well. Anecdotally, subscribers become desensitized to persistent donation requests in political emails [14].

How do senders attempt to overcome fundraising fatigue? We observed many strategies. The most common tactic is to mention a topical hook in the subject line and begin the email discussing that topic before pivoting to a donation request. In a random sample of 100 emails we examined, only 4 of subject lines explicitly mentioned fundraising. To quantify the prevalence of the “pivot to fundraising” pattern, we measure the extent to which sentences in various locations in an email relate to its subject line, as well as the probability that they are explicitly about fundraising (Figure 3). We see that emails don’t stick to the topic of their
Manipulative ways to nudge recipients to open emails

Whether the goal of an email is fundraising or something else, the sender’s first challenge is to convince the recipient to open it. Indeed, the typical open rate for political emails is reportedly only about 20% \[20\] [21]. We found two main ways in which senders attempt to manipulate recipients into opening emails: by employing various types of clickbait and by exploiting the email user interface.

Clickbait resists a universal definition, but broadly speaking it uses the exploitation of cognitive vulnerability to encourage users to interact with content. We analyze three specific types identified in prior work:
• **Forward referencing** or information withholding, whose effectiveness is often explained with reference to the information-gap theory of curiosity \[22\]. We used Blom and Hansen’s definition \[23\] and the 8 subcategories defined by them as a guide to identifying this pattern. (Examples in our corpus: “bumping this for you”; “let’s prove him wrong”.)

• **Sensationalism**, which is a style that “triggers emotion for the reader and treats an issue in a predominantly tabloid-like way” \[24\]. It creates exaggeration and shock, often with appeals to group identities. In practice, it is characterized by liberal use of capitalization, punctuation, and emoji. (Examples in our corpus: “(no!) Mark Kelly SLANDERED!” and “HUGE ANNOUNCEMENT”)

• **Urgency** or time scarcity, which may lead to inaccurate decision making \[25\], emotional exhaustion \[26\], and neglecting other concerns \[27\]. In our corpus, this usually takes the form of deadlines for fundraising or contests. (Examples in our corpus: “April Deadline (via Team Graham)”; “1 huge goal, 1 last chance to help reach it!”)

Turning to user interface manipulation, we again identified three types. We note the parallels with commercial dark patterns, which are “user interfaces that benefit an online service by leading users into making decisions they might not otherwise make.” \[28, 29\]

• **Obscured name**. The “From” field of an email indicates the identity of the sender to the recipient. Some senders obscured their identity, making it impossible for the recipient to learn who sent the email without opening it first. Although the email address is not obscured in these emails, most email clients don’t display that information before opening the email. (Examples in our corpus: “Articles of Impeachment”; “INCOMING: Trump’s REVENGE”)

• **Ongoing thread**. Many email clients display groups of emails as threads. For instance, if the recipient replied to an email sent by Tedra Cobb, a U.S. House candidate, it might be displayed as Tedra, me (2) indicating that there are two emails in the thread. But the From field above — a real example from our corpus — was in fact an isolated email from the Cobb campaign and not part of a thread. Many senders formatted the From field as above to deceive recipients into believing they had received a new message in an ongoing conversation. (Example in our corpus: “John, me (2)”)

• **Re: / Fwd:**. Another marker of an ongoing conversation is a “Re:” or “Fwd:” prefix to the Subject field, indicating a reply or a forwarded message respectively. However, these become a dark pattern when they are not used in a reply to an email (re:) nor contain a forwarded message (fwd:). Examples in our corpus: “re: Carolyn’s email”; “fwd: Falling short.”)

Figure 4 displays two annotated example emails that showcase several of the tactics. In the email on the left, the number of donations left decreases randomly once the email has been opened in order to prompt the user to feel a sense of urgency, a manipulative tactic not covered under our list here.

To quantify the prevalence of these six manipulative tactics, we used a combination of machine learning and regular expressions. The “ongoing thread” and “Re: / Fwd:” patterns were straightforward to detect using regular expressions. We built random forest classifiers for each of the other four.

Specifically, we first randomly sampled 1,227 emails from our dataset. Two researchers then created a codebook to arrive at a shared understanding of each of the three clickbait strategies and the “obscured name” dark pattern. Both researchers independently labeled 250 of the 1,227 emails, observing the subject lines and the from field where necessary. Cohen’s kappa \((k)\) was greater than 0.9 for each of the four strategies, indicating high agreement. One researcher then labeled the remaining emails in the sample. We performed 5-fold cross validation to pick a classifier. The F1 score on a test set of 400 emails was 0.82 for forward referencing and 0.88 or higher for the other three. We also tested whether the classifiers were calibrated on the test set. If they were not—meaning the predicted probabilities of the classifiers were not aligned with the
true prevalence of these four manipulative tactics—we calibrated them on a separate validation set. The supplementary material contains more information about how we trained the classifiers, the features we used, and their performance.

We scored the entire corpus of emails and aggregated them by “active” senders, that is those who sent at least 10 emails that was not a “welcome” email. This retained 924 senders. We see that manipulative tactics are common (Figure 5). The median sender used at least one manipulative tactic in 43.3% of their emails. Few senders desist such tactics: only 1.1% of senders avoided any manipulative tactics in over 90% of emails. Worryingly, a substantial minority of senders, 10.6%, employ them in over two thirds of their emails. This minority is equal parts campaigns and organizations but the latter dominate the rankings by volume. The top senders in this group include prominent PACs and Super PACs like EDF Action Votes, Latino Victor Fund, and the Progressive Turnout Project. In general, the differences in likelihood of use of manipulation with respect to various sender attributes (party, incumbency, campaign vs organization, and electoral race) are relatively minor. Figure 6.A and Figure 6.B show that the clickbait tactics are prevalent but the outright deceptive patterns (ongoing thread and re: / fwd:) are relatively rare.
Email address sharing between senders

Political campaigns often collaborate and consolidate power by sharing or selling users’ email addresses with each other [30]. However, if this is done without subscribers’ informed consent, it violates their privacy.

During sign-up, our bot generated and used a unique email address on each sender’s website, allowing us to match every email to a sign-up. We manually examined every email that was sent from a different domain name than the sign-up domain and, in each case, determined whether the difference in domain names was because the email had an unexpected sender. We did this by examining the domain’s ownership using a
combination of online search, the website’s privacy policy disclosures, and by querying WHOIS, a central registry of domain ownership.

We observed a total of 348 instances of email sharing by 200 entities: 153 federal and state candidates (6.26% of all candidates that sent us least one email) and 47 organizations (12% of all organizations that sent at least one email). The entities receiving these shared emails include 152 campaigns and 130 political organizations, but also 15 news websites. Our findings reveal that email sharing is more prevalent than previously reported \[30\] and that it involves a wider variety of political and non-political entities beyond federal campaigns and PACs. Our measurements are underestimates of the amount of sharing that occurs in the entire political cycle due to our relatively short observation window of about six months.

For each of the 200 entities that shared email addresses, we attempted to determine whether this fact was disclosed to subscribers by examining campaign websites in August of 2020. The majority (114) had no privacy policy on their websites. Only about a quarter (48) disclosed their practice of email sharing in their privacy policies. Out of the remaining 38, 12 did not mention email sharing in their privacy policies, 14 falsely claimed they do not share any personal information with other entities, and 12 had disclosures about email sharing that were too ambiguous to determine whether or not the sharing we observed was permitted.

**Discussion**

We present and analyze a corpus of over 100,000 political emails from the 2020 U.S. election cycle and demonstrate the prevalence of manipulative practices in political emails. We find many manipulative tactics in the Subject lines and From fields geared towards nudging recipients to open emails they might not otherwise open. Our analysis of the email bodies is less comprehensive but nonetheless reveals a few manipulative approaches to fundraising. Additionally, we uncover cases of senders sharing email addresses amongst themselves, including some whose behavior contradicts their own privacy policies.

Manipulative practices are those “covertly influenc[ing] another person’s decision-making by targeting and exploiting their decision-making vulnerabilities” \[31\]. Manipulative political discourse corrodes the ideal of a public sphere and of a participatory democracy \[32\]. It distorts political outcomes by advantaging those who are skilled at deploying technological tricks, triggering a race to the bottom. The normalization of questionable tactics also poses a security risk, making it easier for scammers to imitate political candidates and organizations in order to steal money from would-be donors \[33\].

Three major underpinnings of online manipulation have been identified: the use of behavioral science to identify ways to trigger and exploit psychological vulnerabilities and irrationalities toward desired outcomes; the use of experiments, such as A/B testing, to operationalize those scientific insights; and data-driven personalization of persuasion \[34\] \[35\]. Political campaigns have adopted all three in their email strategy \[2\] \[3\]. The corpus that we analyze, however, does not allow us to analyze highly-personalized or micro-targeted emails due to the limited information provided during signup and the fact that we used fake personas.

The majority of types of manipulative tactics that we uncover are specific to the email user interface. This is clear in the case of tactics aimed at enticing recipients to open emails, as those have specifically developed around the capabilities and limits of email clients. But we believe that even tactics such as bait-and-switch surveys are especially well-suited to the email medium. We hypothesize that this is because emails appeal to the illusion of personalization to tell readers that their feedback is specifically being sought. In our manual examination we frequently encountered messages telling recipients that they had been hand picked by political leaders for a survey, raffle, or another opportunity. We defer a fuller investigation to future work.

Unfortunately, our data do not allow for a direct examination of the targeting or effectiveness of these manipulative tactics. Users with low digital literacy \[36\]—which is closely proxied by age in U.S. \[37\]—are
likely more susceptible to manipulative messaging and adversarial user interface design. This combined with the strong age moderation effects of exposure to digital political misinformation [38] suggests that older Americans are particularly at risk of these tactics. 

There is also abundant evidence from campaign staff that email tactics are subject to careful study, suggesting that the tactics that we describe are indeed effective for fundraising and other political goals [14, 39]. Campaigns likely make a calculated assessment that the gains from such tactics outweigh the losses from unsubscription. There is even the possibility that the self-selected audience becomes more skewed over time, justifying more aggressive tactics—an adversarial feedback loop not unlike the dynamics of clickbait in media more broadly [40].

Another factor that could explain the high prevalence of manipulative tactics is the unfiltered nature of email compared to other media such as online ads. Indeed, as social media platforms and online ad platforms impose fact checking and other limits on political speech, it is possible that some disinformation and manipulative content will migrate to email. This suggests a potential countermeasure, namely for email providers to analyze political emails to add warnings to the user interface or filter out some emails entirely. However, this violates norms around the privacy of email content and also raises concerns about suppression of political speech by private actors; Gmail’s spam filtering has already been criticized along these lines [41].

We observe a striking convergence by thousands of senders to a small number of manipulative tactics and patterns, similar to a recent study of commercial dark patterns [29, 28]. This may be due to campaigns borrowing ideas from each other or due to the possibility that a small set of cognitive biases are being exploited. These hypotheses can be tested by studies of campaign staff and political email subscribers, respectively.

We hope that our corpus will be useful for studying a wide array of traditional political science questions, including how candidates represent themselves to their would-be constituents, how and when campaigns go negative, and what tactics campaigns and organizations use to raise money and mobilize voters. Emails also provide a unique window into smaller campaigns that lack the funding necessary to purchase airtime and consequently do not appear in major datasets of television ads such as the Wesleyan Media Project. Our collection of emails, available at [electionemails2020.org](http://electionemails2020.org), complements other collections including the campaign emails at [politicailemails.org](http://politicailemails.org) and the [dcinbox.com](http://dcinbox.com) collection of official newsletters from congressional offices [42]. We hope that our dataset can shed some light on this previously underexplored, yet highly prevalent, form of political communication.

Acknowledgements

We are grateful to Mihir Kshirsagar, Matthew Salganik, Andy Guess and Orestis Papakyriakopoulos for feedback on our paper. We also thank Justin Grimmer and Jonathan Mayer for advice at earlier stages of the project. We are also grateful to Ballotpedia and OpenSecrets for providing and making their data available to us. This project was funded in part by generous support from the Data Driven Social Science Initiative at Princeton University.

References


[33] Maggie Severns and Scott Bland. ‘scam pacs’ rake in millions under guise of charity. 


[41] Adrianne Jeffries and Leon Yin. To gmail, most black lives matter emails are “promotions”.

1 Additional details about the email corpus

Table 1: Breakdown of the entities in our corpus.

<table>
<thead>
<tr>
<th>Category</th>
<th>Total</th>
<th>Website present</th>
<th>Sent us at least one email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal candidates</td>
<td>4,195</td>
<td>2,548</td>
<td>1,084</td>
</tr>
<tr>
<td>State candidates</td>
<td>9,028</td>
<td>5,536</td>
<td>1,359</td>
</tr>
<tr>
<td>Leadership/Single-issue PAC</td>
<td>679</td>
<td>253</td>
<td>85</td>
</tr>
<tr>
<td>Super PAC</td>
<td>1,721</td>
<td>871</td>
<td>131</td>
</tr>
<tr>
<td>Hybrid PAC</td>
<td>162</td>
<td>139</td>
<td>48</td>
</tr>
<tr>
<td>Other 527 groups</td>
<td>73</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>Other orgs</td>
<td>204</td>
<td>201</td>
<td>100</td>
</tr>
</tbody>
</table>
2 Ethical considerations

Our study was approved by the Institutional Review Board (IRB) of Princeton University. Here we describe some of our research design choices and the ethical considerations that went into them.

We created a fictional profile of a member of the public and signaled to the campaigns and organizations that we were interested in receiving their emails. As a result, we did not collect their informed consent nor did we debrief them about our study. We reasoned that this was necessary because informing campaigns about our data collection and analyses might cause them to alter their behavior or block us from observing.

We considered the possibility that campaigns and organizations might have wasted their time and resources into sending emails that resulted in no action or response. We concluded that the probability and magnitude of potential harm is minimal since email blasts are sent in an automated way.

Beyond our interactions with signing up on websites, we did not donate money to campaigns nor organizations. We did not enter any personal information belonging to a living individual on the websites. We used a rather common first and last name, a unique email address, and an invalid phone number.
3 Topic modeling

This section includes additional information about the Structural Topic Model used to describe content in the email bodies of our corpus. 1,027 out of the 106,342 emails in our corpus did not include any body texts and therefore were removed for topic modeling. We used a variety of R packages to conduct analysis included in the main paper and this supplementary document [1, 2, 3, 4, 5, 6, 7, 8, 9, 10].

3.1 Topics and labels

Table 2 provides an overview of the 65 topics in our model. It includes:

- Topic Label: the label we assigned to each topic during our validation procedure
- Label Category: an additional categorization for each topic as discussed in the main paper
- Proportion (top 15): the overall proportion for each topic in our corpus.
- FREX terms (top 15): terms that are frequent as well as exclusive for a given topic [11]
- Probability terms: terms with the highest likelihood for a given topic

<table>
<thead>
<tr>
<th>Topic Label</th>
<th>Label Category</th>
<th>Prop.</th>
<th>FREX Terms</th>
<th>Probability Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Democratic Campaigns a. Events</td>
<td>Political</td>
<td>0.013</td>
<td>train, imagine, national_democratic, training_committee, opt, every, clock, listen, honest, work_around, promise, incredibly, reason, win, democrat, democrats</td>
<td>every, train, support, imagine, democrats, win, candidate, national_democratic, training_committee, opt, listen, can, know, clock, democrat, honest</td>
</tr>
<tr>
<td>healthcare</td>
<td>Political Issues</td>
<td>0.018</td>
<td>health_care, healthcare, affordable, care, access, drug, cost, insurance, prescription, fight, coverage, medicare, corporate, housing, work_family</td>
<td>fight, health_care, people, healthcare, need, care, affordable, access, can, work, americans, congress, cost, family, drug</td>
</tr>
<tr>
<td>Category</td>
<td>Subcategory</td>
<td>TF-IDF Score</td>
<td>Keywords</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>ordering books</td>
<td>Misc. / Other</td>
<td>0.006</td>
<td>book, check, fill, order, p.s., copy, anyone, form, zero, download, dan, sale, purchase, browser, prepare</td>
<td></td>
</tr>
<tr>
<td>foreign affairs</td>
<td>Political Issues</td>
<td>0.011</td>
<td>military, china, war, veteran, iran, chinese, trade, israel, jewish, security, peace, communist, matt, immigration, united_states</td>
<td></td>
</tr>
<tr>
<td>surveys a. questions</td>
<td>Misc. / Other</td>
<td>0.007</td>
<td>question, list, answer, yes, link, please, ask, simply, contact, submit, subscribe, delete, mistake, confirmation, hesitate</td>
<td></td>
</tr>
<tr>
<td>anti-Trump</td>
<td>Political Events</td>
<td>0.021</td>
<td>trump, donald_trump, lie, attack, obama, adam, schiff, presidency, rally, moveon, white_house, defeat_trump, tell, hate, away</td>
<td></td>
</tr>
<tr>
<td>holiday wishes a. birthdays</td>
<td>Ceremonial / Niceties</td>
<td>0.017</td>
<td>right_reserve, birthday, copyright_©, card, celebrate, mailing_address, wish, love, holiday, happy, day, note, want, enjoy, great</td>
<td></td>
</tr>
<tr>
<td>asking for support</td>
<td>Political Events</td>
<td>0.011</td>
<td>please_click, democratic, elect_democrats, agenda, country, republicans, democrats, across, ameripac, stop, project, elect, dangerous, donald_trump, support</td>
<td></td>
</tr>
<tr>
<td>emails a.</td>
<td>Ceremonial / Niceties</td>
<td>0.016</td>
<td>receive, email, update, send, can_click, change, stay, like, believe, touch, via, good_way, want, support, stop</td>
<td>email, receive, like, send, update, change, support, stay, campaign, want, can_click, make, believe, stop, touch</td>
</tr>
<tr>
<td>news a.</td>
<td>Misc. / Other</td>
<td>0.008</td>
<td>fox, news, tucker, kim, medium, bad, air, msnbc, radio, press, television, cnn, jim, truth, tv</td>
<td>news, bad, medium, fox, air, show, report, truth, press, tucker, kim, watch, week, tv, lie</td>
</tr>
<tr>
<td>education</td>
<td>Political Issues</td>
<td>0.007</td>
<td>child, school, student, education, teacher, public, parent, animal, taxpayer, college, kid, graduate, educator, culture, loop</td>
<td>public, school, education, child, student, teacher, college, high, parent, taxpayer, fund, government, year, animal, get</td>
</tr>
<tr>
<td>guns</td>
<td>Political Issues</td>
<td>0.018</td>
<td>end_citizens_united, nra, brady_pac, gun, gun_violence, authorized, entire, paid, pass_sweeping campaign_finance_reform, end_citizens_united_pac, gun_safety, team_thank, life, ad, wonder</td>
<td>candidate, gun, support, end_citizens_united, entire, gun_violence, nra, life, brady_pac, pass, alex, paid, lose, ad, keep</td>
</tr>
<tr>
<td>supporting Joe Biden</td>
<td>Political Campaigns a. Events</td>
<td>0.013</td>
<td>file, joe, soul, joe_biden, elect_joe_biden, much, thank, sorry, nation, demand, donate, history, personally, click, change</td>
<td>thank, email, support, click, donate, can, campaign, file, joe, much, joe_biden, change, demand, nation, country</td>
</tr>
<tr>
<td>misc 4</td>
<td>Misc. / Other</td>
<td>0.005</td>
<td>new_york, fauci, revenuestripe, tom, dr., times, anthony, jackie, pennsylvania, central, donna, learn, coast, star, scott</td>
<td>new_york, dr., tom, learn, fauci, revenuestripe, pennsylvania, times, anthony, central, fire, john, coast, scott, star</td>
</tr>
<tr>
<td>winning races a. Voter Mobilization</td>
<td>0.026</td>
<td>seat, flip, arizona, blue, race, red, house, gop, take_back, senate, republican, senate_seat, minnesota, win, battleground</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pro-NRCC Political Campaigns a. Events</td>
<td>0.011</td>
<td>nrcc, national_republican_congressional_committee, please_visit, reply, number, phone, text, conservatives, Nunes, correct, devin, comment, u.s._house, prefer, exclusively</td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivation, asking for support a. Explicit Fundraising</td>
<td>0.057</td>
<td>build, go, together, right_now, mean, strong, campaign, today, win, month, lot, november, folk, ask, make</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey politics Political Campaigns a. Events</td>
<td>0.007</td>
<td>jersey, new, josh, elise, brand, stefanik, teresa, south, van, northern, jeff, drew, new_mexico, kean, home</td>
<td></td>
<td></td>
</tr>
<tr>
<td>misc 2 Misc. / Other</td>
<td>0.012</td>
<td>census, chicago, illinois, site, indiana, application, available, reopen, business, ward, covid-19, apply, assistance, relate, testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>misc 3 Misc. / Other</td>
<td>0.007</td>
<td>ignore, confirm, florida, unless, jr., warn, respond, steal, jeopardy, try, debbie, otherwise, newt, karl, final</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Subcategory</td>
<td>Frequency</td>
<td>Terms</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Democrats</td>
<td>Political</td>
<td>0.017</td>
<td>mcconnell, mitch_mcconnell, susan, collins, kentucky, tie, amy_mcgrath, triple, poll_show, defeat, mitch, senate, senate_majority, match, sara</td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td>Political</td>
<td>0.007</td>
<td>trust, american, dream, free, sticker, immigrant, shirt, claim, constant, limit, today, flag, oz, limited, edition</td>
<td></td>
</tr>
<tr>
<td>Campaigning</td>
<td>Events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saving postal</td>
<td>Political</td>
<td>0.007</td>
<td>postal_service, usps, post_office, save, enable, funding, emergency, shut, collapse, postal, united_states, destroy, bankrupt, mail, u.s.</td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td>Issues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGBTQ</td>
<td>Political</td>
<td>0.01</td>
<td>lgbtq, equality, lgbtq+, victory, gay, president_obama, openly, pro, candidate, elect, pac, november, local, unlike, towards</td>
<td></td>
</tr>
<tr>
<td>Signing petitions</td>
<td>Misc. / Other</td>
<td>0.011</td>
<td>add, sign, petition, name, signature, demand, americans, collect, action, call, agree, stand, resign, american_people, put</td>
<td></td>
</tr>
<tr>
<td>Primary debates</td>
<td>Political</td>
<td>0.015</td>
<td>debate, iowa, super, primary, new_hampshire, tuesday, cory, bernie, nominee, pete, stage, caucus, nomination, bernie_sanders, presidential</td>
<td></td>
</tr>
<tr>
<td>Campaigns</td>
<td>Events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retaking Senate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>youth organizations</td>
<td>Political Campaigns a. Events</td>
<td>0.013</td>
<td>organize, young, tool, coalition, foundation, organizer, ed, partner, organization, engage, activist, local, youth, advocacy, indivisible</td>
<td>work, community, member, action, movement, local, support, organization, organize, group, build, read, policy, young, progressive</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
<td>-------</td>
<td>---------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
</tr>
<tr>
<td>Virgina politics</td>
<td>Political Campaigns a. Events</td>
<td>0.007</td>
<td>virginia, northam, assembly, richmond, ralph, gov., county, commonwealth, accord, city, virginians, vpap, board, monday, friday</td>
<td>virginia, county, say, state, city, year, assembly, new, two, gov., general, northam, one, richmond, report</td>
</tr>
<tr>
<td>presidential</td>
<td>Political Issues</td>
<td>0.013</td>
<td>impeachment, trial, impeach, president, witness, evidence, remove, oath, democracy, american_people, constitution, abuse, article, sham, office</td>
<td>president, impeachment, president_trump, trial, office, democracy, house, vote, democrats, senate, power, american_people, senator, can, impeach</td>
</tr>
<tr>
<td>government politics</td>
<td>Political Campaigns a. Events</td>
<td>0.02</td>
<td>washington, politician, politic, real, political, common, problem, put, government, sense, d.c., back, solution, pledge, job</td>
<td>washington, year, put, people, get, political, work, back, real, politic, government, time, job, politician, like</td>
</tr>
<tr>
<td>donations</td>
<td>Explicit Fundraising</td>
<td>0.035</td>
<td>actblue_expression, ad, immediately, payment_information, opponent, donation, money, spend, save, dollar, dark_money, go, chip, million, special_interest</td>
<td>donation, go, immediately, ad, actblue_expression, save, help, opponent, money, payment_information, can, million, chip, campaign, spend</td>
</tr>
<tr>
<td>family</td>
<td>Political Issues</td>
<td>0.032</td>
<td>story, feel, life, mother, heart, something, think, many, different, family, someone, always, well, little, faith</td>
<td>life, know, one, family, time, many, people, see, well, say, think, can, good, like, make</td>
</tr>
</tbody>
</table>
radical left Political Issues 0.021 conservative, nancy_pelosi, liberal, socialist, radical, left, aoc, pelosi, president_trump, radical_left, leftist, socialism, agenda, opponent, border conservative, president_trump, help, liberal, democrats, support, nancy_pelosi, stand, radical, need, fight, socialist, left, can, agenda

passing bills Political Campaigns a. Events 0.011 bill, act, legislation, pass, introduce, bipartisan, co, house, block, reform, measure, colleague, citizen, senator, law bill, pass, act, house, legislation, congress, senate, americans, introduce, law, senator, support, protect, now, ensure

surveys Political Campaigns a. Events 0.026 survey, input, response, poll, strategy, select, joe_biden, diverse, respond, biden, vice_president, minute, presidential, complete, datum poll, response, democrats, survey, joe_biden, democratic, strategy, take, respond, select, input, thank, just, diverse, need

Trump, MAGA Political Campaigns a. Events 0.015 membership, trump_make america_great, communities, 11:59 pm tonight, donald, president_trump, renew, j., schools, make_america_great, official, president, patriot, committee, regard president_trump, trump, committee, president, first, membership, official, team, list, need, trump_make america_great, name, know, see, support

economic concerns Political Issues 0.013 small_business, benefit, tax, relief, billion, economic, worker, unemployment, loan, program, cut, business, income, paycheck, budget small_business, benefit, pay, worker, tax, program, relief, million, economic, cut, business, billion, job, economy, fund
fundraising
a. deadlines
Explicit
Fundraising
0.051
goal, hit, deadline, end, fundraising_goal, quarter, midnight, month, reach, raise, short, fundraising, close, tomorrow, fundraising_deadline
goal, end, help, can, hit, reach, deadline, just, raise, donation, need, month, midnight, fundraising, close

voting
Explicit
Voter Mobilization
0.032
volunteer, voter, door, phone, early, primary, weekend, week, knock, field, march, election_day, talk, reach, team
help, voter, campaign, get, volunteer, need, can, team, week, primary, reach, call, day, message, phone

fundraising contributions
Explicit
Fundraising
0.008
house_majority_pac, contribution_account, per, non, unlimited, federal, may, contribute, individual, account, contribution, president_obama, formula, allocation, labor
contribution, may, house_majority_pac, contribution_account, contribute, federal, non, per, president_obama, individual, agree, need, amount, account, candidate

staying safe
Political Issues
0.016
digital, stay_safe, online, person, cancel, able, hope, everyone, healthy, difficult_time, ever, continue, event, team, time
campaign, time, digital, hope, online, team, able, work, person, continue, everyone, ever, keep, stay_safe, important

climate change a.
environment
Political Issues
0.01
climate, environmental, environment, water, climate_change, planet, clean, energy, scientist, wildlife, earth, green, oil, science, land
climate, climate_change, protect, water, energy, environmental, clean, environment, green, crisis, action, land, science, planet, future

congresspeople, representatives
Political Campaigns a.
Events
0.025
district, represent, representative, congressional_district, congress, proud, congressman, endorsement, serve, community, run, leader, bring, carolyn, endorse
congress, district, run, community, fight, represent, campaign, support, leader, representative, elect, serve, work, proud, first
<table>
<thead>
<tr>
<th>progressive voter turnout</th>
<th>Political Campaigns a. Events</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>democratic voter turnout, turnout, medicare, progress, can_make, high, glad, project, path, mission, matter, every_single, drug, voter_turnout can, progressive, medicare, candidate, turnout, democratic voter_turnout, just, high, thank, way, progress, election, help, fight, democrats</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>social media Ceremonial / Niceties</th>
<th>0.016</th>
</tr>
</thead>
<tbody>
<tr>
<td>share, facebook, video, follow, twitter, social_medium, click, page, post, watch, website, information, friend, instagram, please click, share, can, please, follow, facebook, thank, friend, video, information, twitter, get, watch, website, social_medium</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>donate to Republicans Explicit Fundraising</th>
<th>0.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>nrsc, ally, senate_majority, 5x, charitable_contribution, federal_income tax_purpose, deductible, defend, president_trump, schumer, senate_republicans, liberal, membership, authorize, takeover president_trump, candidate, democrats, nrsc, ally, senate, senate_majority, defend, 5x, contribution, need, make, help, authorize, pay</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>covid-19 Political Issues</th>
<th>0.024</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronavirus, covid-19, pandemic, virus, crisis, health, outbreak, public_health, worker, medical, cdc, hospital, sick, risk, safe covid-19, coronavirus, crisis, health, pandemic, need, community, help, can, worker, people, home, safe, keep, virus</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>misc Misc. / Other</th>
<th>0.014</th>
</tr>
</thead>
<tbody>
<tr>
<td>state, missouri, level, governor, democratic_party, north_carolina, legislative, map, leadership, secretary, legislature, office, legislator, process, ohio state, governor, election, level, republicans, across, office, leadership, make, missouri, democratic_party, candidate, north_carolina, elect, process</td>
<td></td>
</tr>
</tbody>
</table>
California politics
Political Campaigns a. Events
instantly, info, california, process, christy, smith, longer, store, concern, hear, inform, less, street, express, issue
california, email, instantly, process, info, get, keep, click, want, longer, hear, critical, store, concern, important

Republicans losing Senate a. Events
chuck_schumer, mitch, fellow_conservative, matching, conservative, republican_senate_majority, liberal, alert, majority, senate, seat, reply_help, help_stop, conditions, kentucky
match, conservative, republican, senator_majority, liberal, alert, majority, senate, seat, reply_help, help_stop, conditions

justice, courts a. law
Political Issues
court, barr, attorney, ag, ginsburg, wisconsin, supreme_court, rule, judge, sexual, general, assault, justice, legal, lawsuit
right, justice, law, supreme_court, court, general, attorney, wisconsin, protect, rule, judge, decision, case, legal, barr

democratic senatorial campaign committee a. donations
Explicit Fundraising
dccc, gift, tax_deductible, authorize, committee_solely_dedicated, support_democrats, contribution, work_rely, u.s._senate, dccc_[po_box washington_dcc], candidate, grassroots_supporter_like, pay, committee, 11:59
candidate, contribution, dccc, gift, pay, authorize, committee, tax_deductible, make, support, u.s._senate, like, grassroots_supporter_like, support_democrats, committee_solely_dedicated

michigan senate election
Political Campaigns a. Events
michigan, gary, doug, gretchen, jones, whitmer, alabama, peters, take_back, james, senate, extreme, growth, dan, work_hard
michigan, senate, gary, doug, stand, take_back, win, jones, protect, alabama, gretchen, right, whitmer, time, family
<table>
<thead>
<tr>
<th>Topic</th>
<th>Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women rights a. abortion</td>
<td>Political Issues</td>
<td>0.011</td>
</tr>
<tr>
<td>Teaparty footer</td>
<td>Ceremonial / Niceties</td>
<td>0.016</td>
</tr>
<tr>
<td>Lindsey Graham a. South Carolina</td>
<td>Political Campaigns a. Events</td>
<td>0.011</td>
</tr>
<tr>
<td>Voting 2</td>
<td>Voter Mobilization</td>
<td>0.023</td>
</tr>
<tr>
<td>Winning contests</td>
<td>Misc. / Other</td>
<td>0.004</td>
</tr>
<tr>
<td>BLM, police a. racism</td>
<td>Political Issues</td>
<td>0.014</td>
</tr>
</tbody>
</table>

**Topics:**
- Women rights
- Abortion
- Political issues
- Women's choice
- Abortion pro
- Reproductive rights
- Roe equal
- Democratic effort
- Amendment
- Parenthood
- Teaparty footer
- Ceremonial / Niceties
- Lindsey Graham a. South Carolina
- Voting 2
- Voter Mobilization
- Winning contests
- BLM, police a. racism

**Keywords:**
- Woman
- Choice
- Abortion
- Pro
- Reproductive
- Equal
- Roe
- Democratic
- Amendment
- Parenthood
- Teaparty
- Content
- Publisher
- Advice
- Professional
- Tea party
- Coin
- Publish
- Patriot
- Website
- Research
- Safeguard
- United
- Advertiser
- Privacy policy
- Use
- Lindsey Graham
- Jaime
- Ray
- Ben
- Lindsey
- Part
- Perdue
- New Mexico
- Nancy
- Harrison
- Integral
- U.S. Senate
- Movement
- Vote
- Voting
- Ballot
- Cast
- Voter
- Mail
- Absentee
- Registration
- Election
- Voting right
- Suppression
- Register
- Polling
- Let America vote
- Safely
- Enter
- Winner
- Entry
- Utah
- Promotion
- Prize
- Text
- Resident
- Age
- Trip
- Eastern
- Puerto
- Rules
- Retail
- Columbia
- Police
- Black
- Floyd
- Racism
- Protest
- Racial
- George
- Officer
- White
- Violence
- Murder
- Systemic
- Color
- King
- Injustice
- Black
- Police
- Justice
- Community
- People
- Country
- Must
- Protest
- Racism
- George
- Change
- Violence
- Officer
- Floyd
- Racial
need for grassroots support  Misc. / Other  0.01  alex, power, re_election, sean, message, critical, supporter_like, grassroots_support, jon, far, ruiz, |_change, send, right, email_address  alex, power, message, campaign, help, critical, send, like, thank, re_election, right, fight, donate, sean, supporter_like

3.2 Additional prevalence plots

In the main paper, we showed topic prevalence plots by party for three topics that have been prominent in the national conversation during the observation period are seen as politically polarized. Here, we include plots for additional political issues of interest (Figure 1). The topic “staying safe & healthy” is strongly related to the COVID-19 topic we show in the main paper, but in comparison shows stronger differences across party lines. In general, the topics shown in the figure show differences along the expected line, such as a stronger focus on foreign affairs for Republicans and more content about education by Democrats. The topic guns was discussed more by Democrats with mostly negative connotations, for instance demanding more restrictive gun laws in the U.S.

3.3 Topics and manipulative tactics

In this section, we analyze how and whether the topics associate with the use of any of the manipulative tactics we identified. We do so by computing spearman correlations between topic categories and manipulative scores for each email while differentiating between campaign senders and organization senders. As can be seen in Figure 2, correlations between most topic categories and the use of manipulative tactics are rather weak. There are some exceptions, for instance the negative correlations between the topic category “Ceremonial/Niceties”—which includes acknowledgments and birthday wishes—and deceptive from fields, forward referencing as well as sensationalism. Furthermore, for organizations as well as campaigns, the practice of forward referencing is positively correlated with explicit fundraising topics.
Figure 1: List of topics and their prevalence over time. Ribbons represent 95% uncertainty intervals. Emails of senders with unknown party affiliations are not included.
Figure 2: Spearman correlations between topic categories and dark pattern scores by sender types. First row: organizations such as PACs. Second row: political campaigns.
4 Approaches to fundraising fatigue

This section provides additional details about the “Approaches to Fundraising Fatigue” section in the main paper.

To measure semantic similarity between the email body content and the subject, we used the Universal Sentence Encoder (USE) [12] model. This model embeds each sentence into a vector space that captures their semantic meaning. In this space, computing the cosine similarity score between two embedded sentences allows us to measure their semantic similarity. To quantify how related a sentence from an email is to its subject in an interpretable way, we defined a new probability measure:

\[ P[\text{sim}(\text{sentence, subject}) > \text{sim}(\text{sentence, subject}')] \]

Here, \( \text{subject} \) is the corresponding subject to the \( \text{sentence} \) being measured, \( \text{subject}' \) is a different subject line from the same sender, and \( \text{sim} \) is the similarity score discussed. We sampled \( n = 10 \) other subjects in measuring this value, which means we discarded all senders that have sent less than 10 emails in total. This reduced the number of senders we examine from 2,834 to 924.

To determine if sentences are fundraising asks, we hand-labeled 400 randomly selected sentences from the emails in our dataset, creating a 300-100 train-test split, where the base rate of positive examples is 14%. We trained a logistic regression model and observed a test accuracy of 96% with a FNR of 25% and FPR of 0%.

Stratifying by party affiliation, office level, and incumbency, we see little difference in how different groups’ emails deviate from the subject as the email progresses. However, in terms of fundraising asks, we see differences, as shown in Fig. 3. Democrats seem to have a higher proportion of fundraising asks over Republicans, and Federal campaigns over State ones.

Fig. 4 shows an example of an email that pivots from a topical hook into a fundraising ask.
Hi Alex,

I’ve taken on a lot of fights in my life. Today, I wanted to talk about some of the biggest ones and why they matter in this Senate race. I’m also going to ask you for a contribution to my campaign, and I hope this will help explain why. **(If you don’t have time to read all the way through, you can pitch in here.)**

I dreamed of being a pilot ever since watching Star Wars as a little girl. I was often dismissed with claims that “the front is no place for a woman,” but that wasn’t a good enough reason for me to let go of my dream. So, I became a commissioned officer in the Air Force, qualified for the Air National Guard, and went on to serve three tours in Afghanistan as a combat search-and-rescue and medevac pilot.

During a rescue mission, my helicopter was shot down by the Taliban, and I was injured by enemy gunfire. I was honored to have received both the Purple Heart and the Distinguished Flying Cross with Valor for my service.

But my fight wasn’t over. While I could no longer fly because of my injuries, I wanted to keep serving. But the reality was, no matter what I had proven along the way (and who I had proven right) women weren’t allowed to compete for elite military jobs because of the Ground Combat Exclusion Policy.

The rule wasn’t only unjust, but a detriment to military preparedness. So, I took the fight to DC and advocated (members of both parties) to overturn the policy and open up hundreds of thousands of jobs for women in the military.

It was during this fight that I got a firsthand look at how broken Washington is — and it left me with a lot of concerns about the dysfunctional mess we were leaving for our children. And that’s why I decided to take on my latest fight — running for the U.S. Senate.

That’s why I’m asking for your help. Will you pitch in to my campaign so I can keep up the fight?

If you’ve saved your payment information with ActBlue Express, your donation will go through immediately:

Express donate $5 >>
Express donate $10 >>
Express donate $25 >>

---

The dysfunction and blatant obstruction plaguing Washington is not good enough for the people of Texas, and frankly, it’s beneath us as a country. I’m running to break the bottleneck that’s blocking any constructive progress in Washington from coming to reality, and to take this Senate seat from special interest control for the people of Texas.

It’s time for a change, and that’s why I’m fighting — and why I’m asking you to fight with me.

Thanks,
MJ

---

P.S. You’re receiving this email because you’ve shown an interest in helping build a Senate that works for the people and not corporate PACs or special interest donors. If you’d like to opt out, click here.

MJ Hegar was a member of the US Air Force and the Air National Guard. Use of her military rank, job titles, and photographs in uniform does not imply endorsement from the Department of Defense, Air Force, or Air National Guard.

MJ for Texas
P.O. Box 7156
Round Rock, TX 78683

Contributions or gifts to MJ for Texas are not tax-deductible.

Paid for by MJ for Texas

If you believe you received this message in error or wish to no longer receive email from us, please unsubscribe.

---

Figure 4: An example of an email with subject “Welcome to the Fight” that starts out relevant to the title, and then pivots into a fundraising ask.
5 Manipulative tactics to open emails

5.1 Codebook for the manipulative tactics

We documented six types of manipulative tactics—three clickbait types and three dark pattern types. For four out of the six types, two researchers hand-labeled a random sample of subject lines to enable training machine learning classifiers capable of identifying them. We describe this labelling process using the codebook that we created and used:

- **Forward referencing:** We directly employed Blom and Hansen’s [13] eight forward referencing strategies as a guide to identify forward referencing:
  - Demonstrative pronouns. E.g., “These were Chávez’s last words”
  - Personal pronouns. E.g., “He wants to make the national team wear EU clothes”
  - Adverbs. E.g., “Here you can use 4G with iPhone 5”
  - Definite articles. E.g., “In a few seconds the terror bomb explodes”
  - Ellipsis of obligatory arguments. E.g., “Want(s) to arm Syrian rebels”
  - Imperatives with implicit discourse deictic reference. E.g., “See if your bank is at risk of collapsing”
  - Interrogatives referring to an answer given in the full text. E.g., “Do you live in a violent municipality?”
  - General nouns with implicit discourse deictic reference. E.g., “VIDEO: Gigantic baby born in Texas”

- **Sensationalism:** We coded sensationalism as the use of exaggeration and evoking emotion, especially if it appeared in the context of groups (us vs. them). Additional examples of sensationalism from our corpus include: “Democrats Love Gerrymandering” and “Mitch McConnell PRAYING you’ll IGNORE this email”. We examined for the presence of this tactic in the from field as well as the subject of the emails.

- **Urgency:** We coded urgency as the use of explicit (E.g., “3 days to let the world know we’re here to stay” and “Can you support Team Cora before midnight?”) or implicit deadlines (E.g., “Last chance to TRIPLE your gift >>>”) that guided a particular task. We examined for the presence of this tactic in the from field as well as the subject of the emails.
• Obscured name: We coded a from field “obscured” if it did not lead with the name of the sender of the email.

To detect the ongoing thread and re/fwd subject line dark patterns, we created regular expressions:

• Ongoing thread: If the from field matched any of the four regular expressions: 1) `\([0-9]+\)` and ‘,’ 2) ‘, me’ 3) ‘me,’ or 4) ‘alexbrown’ and not ‘response’

• Re:/Fwd: If the subject of an email matched ‘(?i)(re :|.*re :)’ or ‘(?i)(fwd:|.*fwd:)’. If it did, we made two additional checks: 1) Was the email with “re:” a reply to a previous email? If not, then it contained the dark pattern and 2) Did the email with “fwd:” include the string “forwarded message” in its body? If not, it contained the dark pattern.

5.2 Classifiers

We designed machine learning classifiers to detect the presence of the three types of clickbait and one type of dark pattern. For each tactic, we independently trained a support vector machine and random forest classifier using 5-fold cross validation, and picked the best performing classifier. We used the Scikit-learn library [14] to train the classifiers. Table 3 summarizes the results of the classifiers.

Table 3: Summary of the supervised machine learning classifiers we trained to identify various clickbait and dark pattern types in the dataset. Full circles indicates features that were extracted during training. (c) indicates classifiers that were calibrated using a validation set. The classifiers having a gray background were picked to score the entire corpus.
5.3 Aggregating results by sender

Using the classifiers and regular expressions, we computed the proportion of emails of every sender that contain at least one of the six manipulative tactics we identified. Using the probabilities returned by our classifiers and regular expressions, we first computed the following probability score for each email:

\[
P(E) = 1 - P(E') = 1 - P(ON')P(OT')P(RF')P(FR')P(U')P(S')
\]  

(1)

Where \(E\) represents the event that an email contains any manipulative tactic, and \(ON'\) (Obscured name), \(OT'\) (Ongoing thread), \(FR'\) (Forward referencing), \(RF'\) (Re/Fwd), \(U'\) (Urgency), \(S'\) (Sensationalism) represent corresponding events that an email does not contain the tactic. We then computed and reported the mean of this probability score across all the emails for a given sender. This value represents the proportion of emails of a sender that contain at least one manipulative tactic. For the analysis that we report in the main paper, we only analyzed those senders who had sent at least 10 emails (excluding confirmation emails).
References


