

Should the Powerful Be Moderated? Examining Social Media Users' Perspectives on Moderating Politicians' Inappropriate Speech

ALYVIA WALTERS, Rutgers University, USA

SHAGUN JHAVER, Rutgers University, USA

Prior HCI research has investigated social media users' perspectives on the efficacy and fairness of content moderation tools, as well as their concerns around the ways that powerful individuals may not be subject to the same "rules" as other users. However, we argue that there is a sub-group of powerful accounts that deserve specialized attention: those of politicians. Through a focus group study, we outline user perspectives on how—or if—politicians' inappropriate online speech should be moderated, and what mental models are influencing these perspectives. We find that participants' concerns about platforms' capitalistic functions, the changed relationship between politics and media, the outsized influence of politicians' speech, and First Amendment considerations all play heavily into their conceptions of the risks and benefits of moderating these powerful accounts. We also examine how participants perceive the benefits and shortcomings of different platform-enacted and user-directed moderation solutions to counteract inappropriate politician speech. Building upon our findings, we offer suggestions on how scholars of law and platform governance can help build design and policy solutions that better address this issue. We also discuss how HCI research can contribute to advancing the debates around it.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: content moderation, governance, censorship, platforms, First Amendment, free speech

ACM Reference Format:

Alyvia Walters and Shagun Jhaver. 2026. Should the Powerful Be Moderated? Examining Social Media Users' Perspectives on Moderating Politicians' Inappropriate Speech. *J. ACM* XX, X, Article XX (X 2026), 33 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 Introduction

Since former U.S. President Barack Obama took office in 2008, there has been increased attention to the role that social media plays in our democratic processes. As the first president to make extensive use of social media accounts for campaigning and public communications [6, 28], Obama marked a turning point in U.S. democracy—one that not everyone believes is for the best. Scholars have long signaled the problematic relationship between increased social media use and increased political polarization [62], but now that social media use has become nearly ubiquitous, it seems we are in a double bind: there are notable benefits to having unprecedented access to elected officials' stances and up-to-the-moment communications. But when we cannot trust politicians to use these immense

Authors' Contact Information: Alyvia Walters, alyvia.walters@rutgers.edu, Rutgers University, New Brunswick, NJ, USA; Shagun Jhaver, shagun.jhaver@rutgers.edu, Rutgers University, New Brunswick, NJ, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 1557-735X/2026/X-ARTXX

<https://doi.org/XXXXXXXX.XXXXXXX>

platforms responsibly—and we cannot trust platforms to act in the best interests of society—how does this affect the health of our democracy? And how do we hold politicians accountable for their inappropriate online speech, if at all?

In the U.S. context, the “public right to know” has come to be the discursive logic behind platforms allowing controversial speech by politicians, and it was famously the underlying reason for Meta’s reinstatement of Donald Trump on their platform in 2023 (with a full lift of his accounts’ stricter content regulations in 2024). In Meta’s words, their main obligation in this reinstatement was to weigh their responsibility “to ‘allow political expression’ against [their] responsibility ‘to avoid serious risks to other human rights.’” [15]

The current orientation toward allowing expansive freedom of speech to politicians, even when this same freedom may not be afforded to others [21, 43], means that inappropriate speech by politicians continues on social networking sites (SNS henceforth). This problem is not limited to just the U.S. or the Western World [87], e.g., media reports have accused Facebook of allowing hate speech and misinformation by Indian politicians belonging to the governing parties to protect its market interests [29, 80]. However, we begin considering this problem by focusing in this article on the U.S. context.

In a regulatory climate which favors allowing free flow of *politicians’* speech—not always political speech, at large—even at potential risk to others, the current sociopolitical moment in the United States provides rich context to our inquiry. Further, many large-scale social media platforms (e.g., Facebook, Instagram, Twitter) have been founded and/or are based in Silicon Valley, and their operating logics are shaped by the U.S. political environment, cultural trends, and media coverage—a fact that is of special importance today. As Americans find themselves in a highly polarized society led by a president who was banned from mainstream social media in 2021 after using it to incite violent insurrection on the U.S. Capitol—but was then reinstated—understanding how U.S. users approach and encounter politicians’ inappropriate online speech is crucial.

Through eight focus group sessions, we sought to outline “regular” U.S.-based social media users’ orientations toward moderating inappropriate politician speech online. Our analysis contributes to theoretical discussions on how to regulate—or not—politicians’ speech online, and the many ways this decision is impacted by free speech values, potential for social harm, and the public’s right to know its politicians’ stances. Specifically, it speaks to HCI and CSCW considerations of enhancing political participation online [5, 35, 36, 77] and examining platform regulation of the powerful (and its discontents) from users’ perspectives [11, 31, 41, 51, 68, 89].

However, we contribute an urgent, unique angle to this extant literature, because we sought SNS users’ views on moderating their own politicians. Thus, these are not only users, they are also constituents—the very citizens who are represented by these politicians, and allegedly have a democratic right to hold them to account. As such, we argue that there is a specific nexus of power when it comes to moderating politicians’ online speech, and it is one that, thus far, is deeply under investigated. While prior HCI research on user perspectives on moderating “the powerful,” at large, provides excellent context for this study [41, 51, 67], there is, we argue, a marked difference between elites and *political* elites. The outsized regulatory, social, and capital power that the latter exert over society and over platforms makes them a special type of “powerful” social media profile which must be considered distinctly from other influential social media personalities—and it presents a sub-genre of user-centered research that is ripe for continued study across varied sociopolitical contexts.

In what follows, we capture the nuanced and flexible understandings that these focus group participants shared on the social risks and benefits of politicians’ expansive—and often inappropriate—online speech. We elucidate the many ways their complete distrust in social media platforms interacts with their understandings of the ways that shifts in the political economy of media have

changed expectations of what politicians' communications should look like. We examine their perceptions of the utility of currently available platform-enacted and user-directed content moderation solutions [53, 90] in addressing politician speech. And we share the difficult tension that they feel between four seemingly irreconcilable goals: moderating inappropriate SNS communications from their politicians, preserving this harmful speech for purposes of accountability, upholding free speech values for all in the U.S., and protecting others from the harm that inappropriate politician speech may cause—both at a personal level, and at a global one.

2 Related Work

2.1 Moderating the Powerful

Much research has been done on the modes, controversies, and effects of deplatforming powerful users on social media—with “powerful,” here, having a broad definition which covers any accounts that garner unusually high rates of engagement [7, 30, 38, 41, 51]. There is no consensus on the threshold at which a user can be considered to have “unusually” high engagement; one prior study on moderating influential accounts has suggested that users with over 10,000 followers can be considered powerful [30], while another suggests 100,000 as the baseline [51].

Rather than focusing on follower count, this study extends the literature on moderating powerful entities by arguing that politicians' social media accounts—regardless of their follower count—hold a particular kind of institutionalized and material social influence that invites specialized investigation apart from that of powerful civilians' accounts. Notably, there certainly are politicians' accounts which hold the dual influence of institutional importance *and* high follower count, as politicians have been increasingly theorized as undergoing a “celebritization” effect at least partially through social media [22, 64, 66]. However, even as their status can afford them these dual influences, the nuance and effectiveness of deplatforming politicians is not widely understood.

Prior content moderation research has suggested that the deplatformization—or “banning”—of norm-violating influencers' profiles does, indeed, decrease online attention toward these accounts and their harmful content [41, 42], at least on mainstream platforms. However, there is significant evidence to suggest that those who are banned from mainstream platforms may “regroup” in alternative, fringe spaces [55, 94]. This migration may be “safer” for the general public but may encourage even more toxic behavior within the fringe group [42, 72]. These studies focused on *politicized* content rather than *politicians'* content, explicitly, which is the focal point of the present research. However, these findings provide important insight into the nuanced risks and benefits of moderating political speech on mainstream platforms.

2.1.1 Politicians and Online Content Moderation. President Donald Trump is the most notable case study for the type of research with which we are most interested in engaging, as his Facebook, Instagram, and Twitter (now X) accounts were deleted in 2021 following the violent insurrection on the U.S. Capitol—an event for which Trump was accused of inciting violence and subsequently impeached (and later, acquitted). Apart from this example, banning politicians is not a common practice [21], which makes it difficult to track its effects in a robust, empirical manner¹. However, researchers have looked at Trump as a case study, and though pulling generalizable findings from what is essentially a case study of one is not possible, these studies do allow us to start considering trends that might surround the deplatformization of politicians as a particular subset of “the powerful”.

Müller and Schwarz [76] found that the overall toxicity of tweets sent by Trump followers dropped by about 25% after his account deletion, and the total number of tweets about Trump

¹Ballotpedia notes only eight other U.S. politicians who have been suspended or banned from social media platforms. See https://ballotpedia.org/Elected_officials_suspended_or_banned_from_social_media_platforms

sent, writ large, dropped as well—indicating a drop in engagement with Trump, as a figure, on the platform. These effects were largest for Republicans and for non-political users, but they also found lesser effects for “second-degree Trump followers,” or users who did not directly follow Trump, but followed someone else who did. This overall drop in toxicity and engagement with Trump-related content following his ban on Twitter indicates that deplatformization of norm-violating politicians may have the ability to reduce online toxicity.² However, as will be discussed shortly, there is active debate on the legal, moral, and democratic utility of fully banning politicians from social media platforms. Given such debate, it is helpful to consider what other moderation tactics may be useful apart from full deplatforming.

In what are often referred to as “soft” content moderation strategies, social networking sites will sometimes choose to deploy censorship measures which do not remove the content, but either make it more difficult to find (often referred to as “shadowbanning” [59]) or mark it as potentially harmful or misleading (usually called “labeling” [4]). Chipidza and Yan [14] investigated the effects of labeling content belonging to prominent individuals, using Donald Trump as a case study. They found that instead of curbing the spread of this labeled content, as would be the hope, there was actually a higher rate of retweeting labeled content than unlabeled content, meaning that the labeled objectionable information likely spread farther than its unlabeled counterparts. Zannettou [98] also found that tweets with warning labels—tweets which were overwhelmingly political in nature—have more engagement than those without, though this study did not deal exclusively with posts created by politicians, as the present one does.

However, not all platforms apply warning labels to politicians’ posts which contain mis- and disinformation. Though regular users’ posts were, until recently, subject to fact-checking on Meta platforms (i.e., Facebook and Instagram), Fisher et al. [26] revealed that Facebook’s official policy was to avoid fact-checking politicians because, in Meta’s words, “by limiting political speech, we would leave people less informed about what their elected officials are saying and leave politicians less accountable for their words.” (Meta, n.d., as qtd. in [26]) Indeed, the concept of “newsworthiness,” or the idea that it is in the public interest to know what politicians are saying online, is often cited as the reason for lack of content moderation on politicians’ accounts, even when similar content may be at risk for censorship if it were posted by an average user [26]. This is of particular concern given that some scholars argue that the spread of misinformation is a core piece of contemporary political campaigns [83].

Considering the above literature, there is clearly no simple way of devising moderation strategies for political figures, and given the sparse literature on the subject, decisions are seemingly made with little input from experts or from the community. Generally, platforms decide to moderate politicians because they are bowing to pressure, whether it be from the media, the public, or lawmakers, but even these decision-making processes are largely opaque [21, 58]. This study hopes to contribute to these processes by providing a user-centered view on the unique situation of moderating politicians online: a user-centered view which will engage with questions of moderation efficacy, stakeholder responsibility, and user hesitancy to censor politicians.

²Much prior research on “toxic” political content [45, 54, 76] uses machine learning tools to identify language that deserves to be moderated. Perspective API, for example, defines toxicity as “a rude, disrespectful, or unreasonable comment that is likely to make someone leave a discussion,” [79] and it trains its model on discussions posted on Wikipedia and The New York Times. However, our approach is different: we are looking specifically at *politicians’* online posts rather than regular users’ discussion-based content, so what counts, to users, as “deserving” moderation may also be different. We operationalize this concept by asking our participants what politician speech they think is “inappropriate” and deserves to be moderated in the current sociocultural landscape.

2.2 The Unsettled Legal Considerations of Politicians and Platforms

Legally, as things stand in 2025, social media companies are private entities which are not beholden to upholding their users' freedom of expression—that is to say, they can choose to moderate speech in whatever way they see fit [88]. Thus, platforms are entitled to require compliance of user speech with norms of appropriate conduct, as outlined by themselves. These expectations are nearly universally communicated in some form of user guidelines, which is a document to which users and adjudicators can turn when deciding what violates the community norms of the online space [34, 54]. Beyond these platform-specific guidelines, if (or when) something illegal happens in that space—defamation, for instance—the social media platforms are not held liable for these posts, thanks to Section 230 of the Communications Decency Act, which labels platforms as “distributors” of content rather than as “publishers” of it. That is, regulating content on their own accord does not suddenly make platforms liable for meeting any standards of effective moderation [32]. This designation means that Facebook, for example, is not at fault for the (potentially illegal) speech of its users [58]. Though the decision in *Lemmon v. Snap* has provided some opening for sidestepping these legal protections “in cases where harm results from something other than third-party user content” [65]—for example, a design flaw in a platform that could lead to harm—Section 230 is still generally understood to serve as the default legal protection for platforms.

Taken together, platforms attempt to regulate the behavior of their users to fit their community guidelines, and they are ultimately a) not beholden to freedom of speech principles, and b) not responsible for illegal speech which occurs on their site. This is the currently-accepted way of understanding the civic responsibilities of social media platforms. However, social networking sites have now become so influential that the utility of this designation is beginning to be questioned by many entities, including users, the media, policymakers, and legal scholars.

2.2.1 Considering the Stakes of Social Networking Sites as Public vs. Private Entities. When it comes to thinking through platforms' power, at large, some legal scholars say that platforms should now be considered “public forums” due to their operationalization as a public square [46], or they argue that labeling platforms as “distributors” may not be entirely accurate given the hand they have in shaping what is promoted or suppressed on their sites [27]. Some also argue that access to social media should at least have minimal constitutional protections, since those without access are unable to “enjoy their fundamental rights to such an enhanced level as has become standard in recent years” [12]; and still others insist that we are missing the mark if we understand social media platforms only through traditional law concerning the rights of private entities instead of understanding them as systems of governance which should have more accountability to their users [58].

However, there is another piece to this argument that revolves around the constitutional constraints of politicians' social media accounts and if they are being used to conduct official government business, of which there is no single, settled law [8, 40]. For example, while some court decisions have found that politicians have the right to block users or delete their comments—since their accounts are “privately owned channels of communication and are not converted to public property by the use of a public official,” (pulled from the *Morgan v. Bevin* decision, as cited in [40])—others have deemed politicians' accounts as being constitutive of government speech/activity and thus unallowed to ban citizens' access to them (i.e., they are not allowed to block users or delete their comments). This question is further problematized by the fact that many politicians manage multiple accounts (e.g., @realDonaldTrump and @POTUS on X are both managed by current President Donald Trump), and demarcating the public vs. private nature of these accounts—if such a separation exists for such publicly accountable, consequential figures—may prove difficult. Though the decisions discussed here deal with the rights of politicians to block constituents from interacting

with their accounts, and not necessarily with what politicians are or are not allowed to “say” on their accounts, these decisions mark the start of reconsidering the public/private split of government officials’ communications on private platforms, and what “counts” as official government business.

We present this information to make a few concepts clear: first, that there is much debate over how to regulate and legally conceptualize of the type of entity that social media platforms are, at large, and as such, what type of speech (or moderation of such speech) can occur there. Second, we need to make clear that there is unsettled law over what type of free speech protections politicians’ accounts are afforded—or not—and if they are (or should be) different from those of private citizens. Finally, we present these arguments to deepen one of this study’s motivations: that politicians, as powerful people whose online speech often has expansive reach, need to be studied specifically and apart from other “powerful” people. Politicians are clearly in a unique legal position which could affect their behaviors and consequences on platforms, and the policy implications for corporate, legal, and/or government regulation of politicians require special consideration—and should be researched as such.

2.3 User Orientations Toward Moderating Powerful Entities

Because platforms currently operate within the legal means to censor speech as they see fit, they make use of a range of content moderation mechanisms to force or encourage user compliance with their stated community norms around appropriate speech. These moderation processes often operate through a blend of human and algorithmic labor to ensure large-scale compliance from a multitude of angles [33, 50, 82]. There are many ways of doing content moderation, and HCI scholars have conceptualized user- and platform-enacted content moderation as useful categorizations of these tools.

For the purposes of this study, user-enacted content moderation can be understood as the tools available to users to either a) personalize the content that they see on their own feeds through tools like blocking an account or indicating content preferences [25, 52, 95], or b) signal problematic content to the platform through tools such as flagging [17, 60] and crowdsourced fact-checking. An example of the latter is Meta’s use of *Community Notes*, in which a group of users are able to submit and review community notes to provide context to posts that may contain misinformation [93].

Platform-enacted content moderation, on the other hand, can be understood as the ways in which platforms deploy human and machine reviewers to delete (or reduce the reach of) content which defies the community guidelines [68, 71, 75]. These moderation processes are allegedly universal, meaning that average users and powerful users would, ostensibly, be subject to the same levels of scrutiny. While there is evidence that this is not the case [10, 21, 67] (e.g., a Wall Street Journal report [44] shows that Facebook has built a system, called “XCheck,” that shields high-profile users from the site’s regular enforcement), the stated goals of most SNS moderation policies indicate that these are universal expectations—but because the power to establish these regulations and implementation thereof is complicated by platforms’ profit-driven goals, this may not always be the reality [61].

In examining users’ attitudes toward moderation of inappropriate politician speech, we draw from prior HCI research that uses the notion of “folk theories” to understand how people make sense of content moderation processes. DeVito et al. [20] define folk theories as “intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems” [p. 3165]. These theories emerge from users’ lived experiences with technologies, evolve over time, and allow for uncertainty and internal inconsistencies in how people grasp complex systems [19, 97]. Previous studies have used folk theory frameworks to examine how users make sense of algorithmically produced social

media news feeds [19, 24, 56] and how they explain instances of unfair or unwarranted content removals [38, 49, 69]. Our work extends this line of research by examining users' folk theories of whether (and why) social media treat (or should treat) politicians' inappropriate speech differently than that of the average user.

Much HCI literature has investigated user perceptions of the fairness and efficacy of moderation tools, revealing mixed views on social media censorship based on many factors including users' political leanings [2, 3, 90], users' experiences being censored by platforms [37, 38, 68], and the type of harm that the post is forwarding [47, 85]. Many studies have particularly noted that users' attitudes toward content moderation interventions are shaped by third-person effects [18], i.e., presumed media effects on other individuals [47, 53, 81]; we also explore how third-person effects shape perceptions of solutions to combat inappropriate politician speech. User identity can also influence views on the usefulness of content moderation, as marginalized users often feel they are unfairly targeted by platforms and experience removal of their content at higher rates than their peers [37–39, 70, 96].

While marginalized users are concerned how their status may afford them fewer protections in the realm of content moderation, the opposite effect is also concerning to users. As outlined below, prior work has been done on users' unease around the influence of powerful figures on social media, and how this influence might impact the ways that they are shielded from community-wide expectations—even while others remain beholden to these standards. Our study contributes to this literature by investigating user perceptions of fairness in moderating a particular sub-set of “the powerful”: politicians.

One way that this concern around the unequal application of moderation has been understood is through the concept of procedural justice, which refers to user expectations that moderation processes are transparently and universally applied, and that appeals to moderation decisions are seriously considered [9, 78, 91, 92]. Users have been found to feel that those with more “power”—that is to say, those with greater influence on social media platforms—are not necessarily beholden to the same moderation practices as others. Ma and Kou [67], for instance, found that monetized influencers on YouTube felt that accounts with larger followings are less often moderated than they—the smaller influencers—are. Similarly, Caplan and Gillespie [10] found that YouTube creators felt that “established media personalities” (p. 6) were seemingly afforded different layers of protection from ad demonetization than their peers. However, it is also the case that some social media users are concerned about culling the reach of any account—influential ones, included—often due to concerns over the First Amendment and freedom of expression [23, 30, 47, 75]. As this study will illustrate, those feelings often extend to protecting the speech of elected politicians, as well.

3 Methods

This study seeks to understand regular social media users' orientations toward moderating politicians' speech on platforms. To achieve this, we conducted focus groups because they present a useful tool for curating group discussion around a topic—especially one that, so far, lacks deep empirical engagement. Focus groups are particularly useful when researchers are interested in not only idea generation—which was long thought of as the main strength of focus groups—but also in how participants, themselves, are comparing their thoughts and experiences with one another and shifting their outlooks on a complex topic [74].

Further, according to Morgan [73], focus groups are particularly useful in helping to answer questions when “there is a power differential between participants and decision makers,” (p. 15) and “when there is a gap between professionals and their target audiences.” (p. 16) Given that there are massive “power differential[s]” and “gap[s]” between social media users and both a) the platforms

they use and b) the politicians with whom this study is concerned, focus groups again emerge as a strong methodological tool in approaching this study.

3.1 Study Settings

Our University's³ Institutional Review Board (IRB) reviewed this study and determined its status as exempt on 22 October 2024.

3.1.1 Participant Recruitment. In order to arrive at these user-centered insights, we sought a diverse group of participants from the XY metropolitan area⁴. Through both physical and digital flyers, word-of-mouth, and Facebook advertising, we recruited interested individuals to participate in a pre-screening survey, which was open October 28, 2024 – March 28, 2025. We received 194 viable responses (383 responses total).

3.1.2 Participant Selection. In this pre-screening survey, we were particularly interested in respondents' pre-existing engagement with politicians' content on social media, and their thoughts on the justness—or lack thereof—in banning persistently problematic politicians from platforms. Based on participants' answers to these questions, we curated eight focus groups of three to six participants each.

The study is not simply about whether users believe that politicians should be allowed to be banned from platforms, but rather it is about the myriad ways that users engage with politicians' inappropriate speech online (and why). Thus, we chose to separate groups based on the following pre-screening survey prompt in order to decrease overtly misaligned values that may have hindered the discussion process:

On social media sites, some politicians engage in harmful rhetoric, and these sites have occasionally decided to ban such speech. However, in the United States, there is a generally-held notion that the public has a right to know about the stances that the government and its elected officials take on different topics.

Given this context, please indicate your level of agreement with the following statement: Social media sites should be able to ban politicians' accounts if their posts are arguably harmful or inappropriate, even if it infringes upon the public's "right to know" about the politicians' stances.

This prompt was followed by seven-point Likert-type response options, ranging from “strongly disagree” to “strongly agree.” Tables 2-9 in Appendix A show participant responses to this question in the “Politician Ban Support?” row. There was then a follow-up short answer response for participants to explain their reasoning. Their responses are displayed under “Short Answer Responses” in Tables 2-9 (Appendix A).

Respondents who indicated any level of agreement were grouped together, and those who indicated any level of disagreement were grouped separately (with “neither agree nor disagree” respondents dispersed between the two). While the goal was to facilitate focus groups of five to eight participants each, session sizes varied based on attendance. A minimum of three participants was required to run the session in order to ensure rich group exchange rather than dyadic conversation. There is an over-representation of Democrats (or those leaning Democrat) in the sample, but it is otherwise fairly diverse. Table 1 presents a summary of our participants' demographic information. Specific information on participants in each of the eight total groups can be found in Tables 2-9 (Appendix A).

³We will specify the University's name after the peer-review process concludes.

⁴We will specify the geographic location after the peer-review process concludes

Race/Ethnicity	#
Black or African American	13
White	14
Mixed Race	3
Gender	
Man	14
Woman	15
Transgender Woman	1
Age	
18-24	2
25-34	4
35-44	3
45-54	9
55-64	6
65+	6
Political Affiliation	
Strong Democrat	10
Not very strong Democrat	2
Independent leaning Democrat	9
Independent	3
Independent leaning Republican	4
Not very strong Republican	0
Strong Republican	2

Table 1. Summary of Participant Demographics. Note: race/ethnicity was a “select all that apply” question format. For purposes of aggregation here, we have collapsed those participants who chose more than one race under a “mixed race” heading. Please see Appendix A for specific racial demographics of mixed-race participants.

3.2 Study Flow

Focus groups ran on a semi-structured protocol, which left the moderator with the freedom to pursue emergent strains of inquiry as the discussion went on. Our questions included how participants defined inappropriate politician speech and how they perceived different ways of moderating politicians (more details included in Appendix B). Each focus group session met for approximately two hours, presenting us with approximately sixteen hours of in-depth conversation as data. Upon completion of the focus group meeting, each participant received a \$50 Amazon gift card as compensation for their time.

3.3 Data Analysis

Using the qualitative coding software Dedoose, we performed a grounded inductive analysis [16] on our 16 hours of data from 30 different participants, defining emergent themes surrounding the moderation of politicians on social media. While close reading all of the transcripts, we allowed codes to emerge, staying close to the data and avoiding pre-conceived coding categories. We initially began by coding the “agree” groups to build out granular codes, then applied them to the “disagree” groups, adding new codes as needed. Examples of codes in this stage included “politicians posting personal attacks” and “use of community notes as moderation.”

After several read-throughs of the eight transcripts, no new granular codes emerged, so we began memo writing and using Dedoose’s analytical affordances to better understand which codes were most often interacting with one another—and why that might be the case. For example, we found that the codes “algorithms” and “(dis)trust in social media” interacted at an unusually high rate, so we began memoing about the other relationships we see between these ideas, and how

we can better position those findings as a piece of the larger story our focus groups were telling us. The authors discussed these emerging interactions, considering the most important trends to define and sharpen.

After this memoing process, we were able to begin a second round of coding, defining higher-order themes and parent codes. Examples of codes at this stage include “defining inappropriate politician speech,” and “changes in media economy.” The authors discussed this consolidation process through several different iterations until a final codebook was agreed upon and applied to each of the eight transcripts, ultimately resulting in the themes elucidated in our findings.

4 Findings

4.1 Ethical Considerations

Our focus group participants made it very clear that their debates on politicians and content moderation were not occurring in a vacuum. Participants’ discussions were, indeed, so guided by the following ethical considerations (sections 4.1.1 and 4.1.2) that they were difficult to code for without highlighting nearly every interaction in the transcripts. Thus, we begin with a delineation of the ethical grounds upon which users are considering questions of online censorship of politicians’ speech.

4.1.1 Content Moderation vs. The Public Right to Know. Though our focus groups were, ostensibly, divided based upon respondents’ beliefs in the rights of social media platforms to ban politicians’ accounts, this opinion very often proved to be flexible. Participants who began the meeting feeling strongly that politicians’ hate speech, for instance, should be removed sometimes ended up questioning whether it is more useful to let politicians’ speech stand as a means of alerting their constituents to their values. P19, for example, stated, “*I want to know if a guy thinks that it's okay to be in the Klan and he's running for office. I want to know that he thinks it's okay [so that I can vote for a different candidate].*”

Apart from making their values known, allowing politicians’ harmful online content to stand came with other arguments, too. According to some participants, the public also has a right to know how false narratives perpetuated by politicians may affect the lives of themselves and their loved ones. P15 pointed out that if all of President Trump’s and Vice President J.D. Vance’s misinformation on immigrants was censored by social networking sites, people might be at risk of not knowing they are in danger of deportation:

So on the one hand you're like, you're tired of [the immigrant-centered vitriol and disinformation]. It's like, it's so terrible, you just want it to go away, but at the same time you want to be aware. Because, you know, next thing you know, you're gonna be hearing a knock at your door.

P19 spoke to one other manifestation of the “public right to know” argument: the potential social dangers of deplatforming inappropriate politicians, entirely:

What happened when they banned Donald Trump from Twitter and Facebook was that Donald Trump, he said things that then only appeared to his audience, and he made himself the victim... by taking him off of Twitter, a lot of people could then not see what an idiot he was... So, it's unpleasant, and I don't like it either. But sometimes the best thing to do is let your enemy tell you who he is.

This nuanced understanding of how politicians’ problematic posts might be more useful to Americans if they remain on social networking platforms—rather than deleting the inappropriate language—was an interesting turn of conversation and speaks to the ethical concerns around how harm can best be avoided. Notably, many participants also mentioned that controversial posts

by politicians are almost never missed by news outlets and other social media users—meaning that they could still come to public attention, even if they were removed via platform moderation mechanisms—but it is useful to recognize the many ways this conversation unfolded, and the contested terrain upon which it stands.

4.1.2 Equal Treatment and Platform Terms of Agreement. Additionally, there was much discourse on moderating normal social media users versus politicians, who have decidedly more cultural influence and are accountable to the citizenry in unique ways. There was obvious tension between the idea that politicians are—or used to be—“held to a higher standard,” (P29) and are now not even necessarily seeing consequences for speech which average citizens are punished for. P18 lamented, “They get away with more than the average person could [on social media],” and was met with general group agreement, until P17 and P19 spoke out, indicating a different point of view which emphasized the magnification of politicians’ speech online:

P17: I would say that it puts the politicians under the microscope that they actually...

P19: That they should be.

P17: Yeah, you understand where I’m going?

Some groups had a similar mix of disagreement, while others were in consensus: politicians are allowed to get away with too much harmful speech online.

Regardless of each group’s sentiment, however, the push and pull between how politicians “should” be held accountable for their social media posts (in comparison to the average user) was an ever-present concern often informed by the notion that politicians are meant to represent their constituents and have extraordinary power and influence. Thus, they should be held to at least the same standard as others, as outlined in platforms’ terms of agreement: “If I am beholden to terms in social media, so should everyone,” asserted P29. She went on to say, “with [politicians’] hate speech, with dangerous speech, with things that are promoting violence and bigotry—those should be taken down because they are against the terms and agreements on these social media platforms.”

4.2 Why Moderate?

These focus groups revolved around how participants felt that politicians’ inappropriate speech should (or should not be) moderated online. In order to come to conclusions on this, we first had to define what “inappropriate” politician speech means to these participants. Each group had a diversity of thoughts on this topic, but they could generally be boiled down to two thematic elements: politician speech can be inappropriate in the ways that anyone’s speech can be inappropriate, like by deploying bigoted ideas, making reductionist assumptions about large groups of people, or spreading misinformation. But there is also a specific way that politician speech is different than the average social media user’s speech, according to many focus group participants: politicians can say something that is inappropriate for their rank and responsibility in society; something that might be alright for others to say, but not for them. As P19 succinctly stated: “Well, I’m allowed to be an asshole. But I think politicians need to have some certain amount of decorum and respect for their constituencies.”

Two overarching themes emerged on why content moderation may be a useful tool when it comes to politicians’ inappropriate speech: first, that the relationship between politics and media has irrevocably changed with the advent of social media, and second, that the new media environment has exponentially expanded the reach and influence of politicians’ speech.

4.2.1 Transformations in the Relationship Between Politics & Media. The story of the changing media economy—and its symbiotic relationship with politics—figured prominently in participants’

understandings of why politicians' speech might be in need of special content moderation practices. "I used to enjoy the thud. The newspaper hitting the door," P2 recalled. He goes on:

As a kid, I used to run to the door to get that newspaper... the Sunday New York Times was my favorite thing to get. And I'm not just saying the tactile feel of it. It was just the way they went into detail on everything. And now everything is about 'I got to get this in; I got to get it all in. And I got two minutes to do it.'

While these narratives may feel like participants waxing nostalgic, it went far beyond that: there was overt understanding that the transition from analog, traditional media to hours of scrolling on social media per day has seismically shifted the relationship between media and politics. The pace, reach, and ease of access of social media has made politicians' thoughts more accessible—and less "checked" (P20)—than ever before. On the pace of social media, P28 said,

We've been on the 24-hour news cycle for quite some time, but they cannot keep up with a 140-character Tweet. Impossible. So sure, there's like talk show hosts and stuff like that, but truly the internet is the place you go to find out what's happening in real time from the politician's mouth, or fingers."

This pace and ease, then, leads to an "erosion" (P8) of the traditional checks and balances we are accustomed to—both in government structures and in the "checks" that traditional media gatekeepers provided. "On the one hand, [politicians' social media use] is incredibly democratic because now a politician can reach directly to his or her constituencies without any kind of filter," P11 stated. He went on: "That, used correctly, would be very democratic. It's a great thing for democracy. Obviously, on the other hand, without somebody as a gatekeeper, they can sell as much bullshit as they feel like it."

Participants often spoke on weakened trust in these traditional gatekeepers, however, noting that capitalism, the explosion of social media, and growing polarization is shooehorning too many news channels into partisan reporting. This ultimately sows public distrust in all types of fact checking, both in traditional media and on social media. Even beyond the "gatekeeper" functionality that journalists have traditionally served, participants expressed a sense that there used to be a degree of separation between politicians' thoughts and their speech: they curated thoughtful responses for public consumption via legacy media outlets. However, these participants overwhelmingly understood some politicians to be posting online with abandon, and sometimes with notable consequences, as will be discussed below.

Though politicians and media have always had a close relationship, one participant deemed the growing ties between politicians and Silicon Valley as "deeply unhealthy" (P28) and riddled with questionable economic gains for both platforms and the politicians, themselves. Another noted that even as politicians' credibility is tied to their online presence—a weak online presence may correlate to being "not even really a serious candidate" (P20)—this expectation of having an active online presence has also fundamentally changed the role of politicians' (online) speech:

I can look at their platforms today. I can look at their platforms next week. And I'm not limited to going to a rally to find out what they think, or [to] go into The League of Women Voters' little pamphlets. I can find it out anytime. And that means that they are always... campaigning, every minute of every day, 24/7. Because people can look at their social media. So that means that their personality becomes more front and center because that's how we're getting our information. And it does make it more personality—cult of personality kind of thing. (P20)

Social media, it seems, is understood to have amplified this "cult of personality," or what some participants spoke on as the celebrityization of politicians. After delineating a history of the slow

intersection of politics and celebrity, P10 said, “*the division between celebrity and politics ultimately went away. And I think that the general American perception is that to be at the highest level of the federal government means you can't just be a traditional politician... You have to be somebody that's talked about, somebody that's in the news, somebody that you turn on, and 'I know that guy.'* Otherwise, you're never going to get the job.” He goes on to explain that this matters because it has “*dramatically altered not just how Americans perceive who they vote for and how they vote, but how they process the news and the information that's coming from their [social media] sources.*”

Taken together, participants narrativized that the shift from analog to social media—a shift which could have been, in theory, great for democracy—has actually sown distrust, encouraged polarization, and made political speech less about communicating trusted information with the public and more about curating an electable personality. These generally negative outcomes of the new media environment create fertile ground, some participants feel, for moderating politicians' speech, as it is, at times, doing more harm, more quickly, with greater reach than ever before.

4.2.2 Outsized Impact of Politicians' Online Speech. Participants asserted that the affordances of the social media environment—rather than those of the traditional media environment—have multiplied the reach of politicians by many times. The implications of this outsized influence was understood in many different ways, from demarcating the differences in the online influence of politicians and that of the average SNS user to reflecting on its dangerous, real-world impacts.

Universally, focus group participants felt that politicians' posts have considerably larger impact than that of the average social media user. P11 explained this sentiment succinctly, saying “*If an average individual posts something [about Haitian immigrants 'eating cats and dogs'⁵], that's bad enough. When someone who is the vice president or the president or a senator that has the bullhorn, who has like, you know—free speech, [it] is totally different. Because my ability to speak and [Vice President] JD Vance's ability to speak are not the same thing at all.*” This, sometimes, was used as reasoning for why moderation may be necessary. In relation to the January 6th organizing⁶, P3 stated that social media is “*very dangerous*” because people can “*organize very effectively through social media, and you can meet people you wouldn't have met otherwise in your neighborhood or who connect with the same strong beliefs and really just feed on each other to take action.*” The political followers, in this case, were contributing to the outsized influence that Trump had as an effect of his inciting social media posts. The “*beer muscles*” (P2) that the anonymity of social media gives users helped create a coalition of people who were willing to unconstitutionally storm the Capitol, but this was led by Trump's rhetoric, which has instantaneous and wide reach through social media.

While the influence of politicians' posts was understood to be vast, participants often noted that they were more concerned about the impact this speech had on others, rather than on themselves. P30 noted that

[politicians] can say whatever they want, but it's not gonna really affect me that much. I'm concerned about the insane, crazy person who's gonna read some of that stuff and do something to hurt somebody based on what they read... So, you know, me blocking it is great, but there needs to be more done than just us individually blocking it.

P1 also noted that there are some “*sick people*” who look up to politicians and take what they say as gospel—even if what they say leads to violent outcomes. These outliers in society, these

⁵In a 2024 presidential debate, Donald Trump falsely claimed that Haitian immigrants were eating their neighbors' pets. His running mate, JD Vance, backed up these claims. See <https://www.bbc.com/news/articles/c77128myezko> for more.

⁶January 6th refers to the 2021 attack on the U.S. Capitol by U.S. citizens. President Donald Trump was ultimately impeached—and later acquitted—for his involvement in inciting these attacks, as he falsely promoted the idea that 2020 election had been “stolen” from him. See <https://www.britannica.com/event/January-6-U-S-Capitol-attack> for overview.

participants feel, must be taken into account when considering what type of politician speech should be allowed to stand.

4.3 How to Moderate Politicians' Speech

Some participants voiced that they were hesitant about any moderation of politicians' speech at all on platforms, barring unconstitutional speech. This was often attributed to politicians' constitutional rights to freedom of expression, and sometimes to the slippery slope that any type of censorship—even by private companies like social networking sites—can lead to: *“If social media platforms are not, you know, attuned to free speech or not a part of free speech, then are we just like the rest of the world where this is not really allowed?”* (P12) Some also spoke on their own right to potentially *“life altering”* (P3) information which could be revealed in politicians' (potentially removed) social media posts, with P3 wondering *“are [platforms] breaking my civil rights by not giving me certain information?”* As difficult as unsavory speech may be to handle on social networking sites, it seems to be a necessary evil for some.

In overt ways, some participants also dabbled in the legal argument surrounding the potential mismatch of the “public square” functionality versus private ownership of social media. P2 argued that the internet should be treated similarly to broadcast airwaves, meaning that it should be publicly “owned” in some capacity. In a colorful analogy, he remarked, *“If [platforms] want to use the public's airwaves, then [platforms] have to give the public complete access to it. You can't be half pregnant. You can't be in and out. It's yes or no.”* Essentially, he is arguing that SNS use of the *“public's [internet] airwaves”* should give the public absolute access to everything published on those *“airwaves”*—without censorship of that material: *“There should be brutal and drastic consequences to companies that use the public airwaves and then want to censor the public.”* (P2)

P12 took a different angle to this argument, reflecting that *“1776 cannot even comprehend Facebook,”* by which she meant that the originators of the Constitution could not possibly predict (and thus, write law for) the ways that speech would travel and be amplified via social media. To that end, she wondered if *“freedom of speech might actually need to be redefined.”* P11 agreed, saying *“1776 has nothing to do with 2025.”* He goes on, in a mocking tone: *“All this happy ‘what did they really mean in 1776’? Well, not a few years before that, they burned witches. So.”*

Many, if not most, participants agreed that there should, perhaps, be more reliable mechanisms in place for moderating politicians' speech that violates the community guidelines of each platform, including the ways in which group members defined politician speech as “inappropriate” through violations such as bigotry, disinformation, and unconstitutional language. Both user-enacted and platform-enacted solutions to this problem surfaced in these focus group discussions.

4.3.1 User-Enacted Moderation. While a few participants indicated use of lesser-known personalized moderation tools, such as changing their content settings to allow “more” or “less” political content, most participants indicated that they simply keep scrolling past politicians' content that they find inappropriate, or they unfollow/block accounts they find unsavory as they appear on their feeds.

The use of community notes was also discussed, given Meta's January 2025 announcement that they intend to transfer to a community notes model of fact checking. Participants had mixed feelings on this, often wondering why they would trust a community note any more than they would trust anything else on SNS. When considering the integration of community notes, P10 said, *“I think it's potentially harmful. That's Wikipedia.”* By this, he was referring to the Wikipedia model of knowledge building—essentially, that anyone can edit the information, and readers must approach it at their own risk. P30 pointed out that politicians' posts, in particular, may become

essentially meaningless with the integration of community notes, as every politician has detractors who may seek to publicly attack them through the community notes functionality.

Others have more hope for this new tool, believing that it may help users pause to consider other perspectives. P4 stated,

assuming that the community notes are trustworthy, at least there's someone out there saying, 'hey, we don't have fact checking as a whole, but here's a collection of people who think differently or who have seen maybe a different article that says it a different way.' And it's like, 'this is not the whole story.' And it prompts people to not take things at face value.

This is supportive of an often-shared sentiment that participants felt users—and themselves—needed to “do their own research” before engaging with politicians' speech.

The Pitfalls of Reliance Upon User-Enacted Moderation Mechanisms

Participants outlined some reasons that they feel personalized moderation (e.g., blocking, muting, changing content settings) may not be enough to combat harmful or inappropriate politician speech. For some users, it is akin to a civic responsibility to use moderation mechanisms, such as flagging, which would demote or censor politicians' content for *everyone*—not only for themselves. P1 thought of flagging as being similar to the “if you see something, say something” adage: *“Even though I know it's not me personally [making change by flagging] on the Internet, at least I feel like I'm doing something. If something happens because I saw it and I didn't say anything... well, you gotta look out for one another, you know?”* P29 shared a similar sentiment, saying, *“for me, I think it is part of your civic duty to hold elected officials accountable. And until, if ever, those elected officials are removed from a social media platform, I consider it my responsibility to, at the very least, try to inform other people and to try to give context where it is needed to things like that.”* The focus on “other people” was a feature of many of these conversations, and suggested that personalized moderation tools, such as blocking or muting accounts, may not be as suitable of a solution due to its effects on only the individual level.

Contrary to these feelings, there were plenty of participants who felt no personal responsibility for moderating politicians' speech, opting for other stakeholders (and the politicians, themselves) to take on that burden. P28 reflected, *“I think, like, 'not my party, not my problem.' I didn't put you [in office]. I'm not going to chase you down... So I think that's why I have so many people blocked, because I just don't want to deal with it.”* In her case, the user-enacted tool of blocking is a suitable fix, because she feels no responsibility for the actions of a politician she did not vote for and did not have a hand in bringing to power.

4.3.2 Platform-Enacted Moderation. After prompting by the moderator, there were several platform-enacted tools which are already in existence (in some form) which some participants agreed may be useful: fact checking, waiting periods (before a politician's post goes live), better-trained AI moderation, larger teams of human moderators, and increased thresholds for account verification for powerful individuals.

We spent additional time interrogating the strengths and pitfalls of fact-checking, as the timing of these focus groups coincided with Meta's announcement of their rollback of third-party fact-checking (and integration of community notes to the platform, as mentioned above). While some participants felt indifferent about slashing third-party fact checking—*“I don't think fact-checking works. I think whoever's doing the fact checking is lying also.”* (P5)—others felt differently. *“I guess my thing is, why do [platforms] want to stop fact-checking?”* (P15)

Apart from the already-existent platform-enacted moderation tools, participants suggested several other platform-based solutions for moderating politicians' speech. One idea was to have

very clear (and publicly transparent) thresholds for deactivation—temporary or permanent—of politicians’ accounts:

In a perfect world where people aren’t just reporting posts because they don’t like them... a politician’s account will be deactivated. Or they couldn’t post if they hit a certain number of reported posts. So if there’s like five posts over a certain amount of time [where] they get reported for being, you know, untrue or offensive or whatever, then they couldn’t post or their account would be deactivated or something. (P4)

Another proposed solution was to remove politicians’ problematic posts, but to then leave a placeholder that indicates why the post was removed:

I would still like [platforms] to flag people. Take their account down for maybe the seven days, like [P30] was saying, but give a reason publicly: “Hey, if you go to X person’s page, they’ve lost the page privilege of this account for making racist remarks, sexist remarks,” you name it. So somebody would be responsible for qualifying them on the social media side, but then the public would say, “oh, this is what happened.” And then you get to come back. But maybe you get like three strikes. (P28)

Interestingly, an idea brought up by several participants was to simply not allow politicians on social media, at all. “I would love if we just kicked all politicians off social media,” P29 said. “It’s not an official platform. Just get rid of them. They don’t need to be there. They have official platforms [referring to their .gov websites]. There’s no reason for them to be on social media.” P22 shared a similar idea, saying “it would be better if all politicians had to post to, you know, socialmedia.gov and there they could share their opinions on socialmedia.gov—outside of the Facebook, Twitter, BlueSky, whatever rules.” The group that P22 was part of mulled over the necessity of having more official, systematic documentation of every social media post made by politicians so that it would become part of the public record of their term in office—just the same as if they posted “a flyer on a tree,” (P23) which would generally be archived as part of their public communications. A different group discussed a similar topic, with P24 stating, “a lot of things online are permanent, but a lot of things aren’t. Or at the very least, because so many things are permanent, it just buries everything...” To address this problem, P24 considered the benefits of having an easily-searchable directory where every politicians’ posts—inclusive of posts deleted by platforms—could be found.

Distrust in Platform Solutions

The overriding, most obvious reason that these focus groups seemed hesitant—or at least indifferent—about most platform-enacted solutions to moderating politicians’ speech is that they, unanimously, do not trust platforms to be good actors. Capitalism played heavily into these conversations, as participants were keenly aware of the alliances between corporations, platforms, and politics. P11 nodded to the unique problem this creates on platforms that claim to increase the democratic power of free speech:

[Is] the platform to make money, or is the platform to promote democracy? Is it to promote free speech or is it to promote profit? There’s a nexus in there that they are not the same thing and they are at odds against each other. [Platforms want us] to do the emotive clicking because that gets eyeballs, that gets advertising, that gets profit. That does not do anything for democratic society. So what are [platforms] after here?

P13 lamented that we have almost nowhere to turn when we want to look for credible information that may allow us to fact check our politicians’ online speech, saying that tech companies have “monopolized” so much of the information landscape—and they are so “compromised” by politics and money—that even the most responsible of media consumers are left with few options. Feelings

of the futility of platform-enacted moderation were rampant, with many participants sharing a similar sentiment: “*you’re up against a losing battle.*” (P3)

Algorithmic opacity was an issue that came up again and again, as well, with participants indicating a distaste for not understanding how or why they were seeing politicians’ content on their social media feeds. There was concern that any engagement with a post—even to denounce it in some way—would ultimately bring more attention to it due to algorithmic processes: “*sometimes, anger or reports can also boost [politicians’ posts] even higher,*” P13 stated. “*So now you think you’re hurting that post, but you end up helping it. There’s just no way to tell.*” The opaque nature of news feed algorithms was an overriding concern, and participants felt that their efforts to influence this algorithm were futile. In a warning call to platforms, P19 mused,

I think that one thing that we need to remember... is that we can’t rely on the platforms to be good actors. That platforms change and they change their agendas. Inevitably, though, these platforms are going to destroy themselves by not regulating this content. You can see the decline of X, or Twitter. If you’re familiar from its heyday—[since then] it’s become almost unusable.

5 Discussion

These focus group discussions make clear that there is no simple solution to the problem of inappropriate politician speech on social media platforms. Even the solutions that platforms *do* try to put forward are largely perceived to be deeply compromised by the economic and partisan politics in which social media platforms have embroiled themselves. Trust in platforms to act toward the public good was almost entirely non-existent in these discussions, and this should be considered an overarching, topline finding: *social media is a tool that could easily foster democratic discussion and increase accessibility to constituents’ politicians, but it is so distrusted by this study’s participants that it is largely perceived to be intentionally sowing discord and destroying democracy to increase its own profits.* But there is much to take away from these focus group meetings, and much gleaned on the public orientation toward freedom of speech and content moderation in the context of U.S. politicians’ online speech.

5.1 Characterizing the Public’s Concerns Around Moderating Politicians’ Speech Online

A nuanced mix of concerns came together to form these focus groups’ opinions on moderating politicians’ online speech. Although not shared by every participant, thematic coding revealed a “story,” so to speak, of how and why everyday users are distrustful of platforms, politicians, and the damage that they can do, together.

5.1.1 Social Media’s Effects on the Political Environment—and Vice Versa. According to this story, *money, politics, and social media have created an unholy alliance that participants are acutely aware of.* The profit-based structures of social media, the power that politicians have to pressure platforms into regulatory action, and the ability for politicians to actively campaign on these platforms—while creating ad revenue and increased engagement for these platforms—has created a media economy which amplifies users’ concerns about inappropriate politician speech. Notably, this was a strongly-shared set of beliefs by participants of all partisan loyalties.

These unprecedented changes in the political economy of media have changed political communication and its rhetorics and functions. With 24/7 access to their constituencies, politicians can post whatever they would like at any time of the day—there is no longer the need to call a press conference, curate a speech, and engage with reporters. Instead, the increased pace and lack of “filter,” so to speak, in political communications has increased the ability for politicians to speak

inappropriately on widely-consumed spaces, such as social media platforms. Focus group participants noted that not only did journalists used to curate a story around the most important points relating to politicians' speech, but also the decorum of the press conference or the "official" public event encouraged a specific type of responsible behavior that has been lost due to the accessibility of social media and the platform logics which reward engagement.

According to these participants, increasing engagement, for politicians, often looks like sensationalism, emotional or intellectual manipulation of voters, or both. In other words, platform affordances have encouraged politicians to seek public engagement in new (often worrying) ways that create a "cult of personality" (P20) around them. This, in turn, amplifies the "celebritization" of politicians, a process theorized to be especially influenced by digital media [22, 64, 66]. Politicians are often now judged based on their personalities, sensationalism, and online engagement—a process which does not necessarily lead to all positive or all negative outcomes, but which does bring an even heavier feeling of "performance" into the political arena for many of these participants, and exponentially increases politicians' audience.

To wrap up the "story," then: this celebritization *does* become problematic when politicians' (online) speech is inappropriate, both for the harm it can cause online and for the ways that their "crazy" (P30) followers may react. President Trump's incitement of the January 6th attacks on the Capitol was conjured again and again—not only by Democrats—as example of the ways that other people can be affected by this now-outsized, instantaneous, and far-reaching influence that politicians have on social media. This contributes to prior HCI and CSCW conceptualizations of online harms and the need for justice-oriented work to address specific vulnerabilities [57, 84–86]. Indeed, it is likely that users' individual experiences of being targeted for their identity or ideological beliefs may sway their folk theories about platform practices and orientations toward moderating politicians, and it is possible that the Democrat lean of our sample impacted how these discussions unraveled. It would be important for future research to investigate how factors such as targeting, marginalization, and political affiliation, more broadly, affect attitudes about this issue.

There was another type of harm spoken on, too. Not only do politicians have tremendous influence on their followers (and on public discourse, in general) through utilizing social media, but they also have outsized influence on the future of the country. *If their sensationalism is encouraged on platforms—and then they can actually do something to follow up on that inappropriate speech due to their sociopolitical power—their inappropriate speech carries a special kind of potential for harm.* For example, participants repeatedly mentioned Trump's false assertion in a 2024 campaign debate that Haitian immigrants were eating their neighbors' pets in Ohio. While a normal user could post this without any real power to *do* anything with it on a large scale, when the president posts this, it is used as justification for his mass deportation policies—deportations which have already led to many illegal actions [13]. This type of power differs from, for example, the highly-popular YouTube personality who breaks the platform's rules but does not get demonetized [67], or the social media influencer who gets away with posting content that her minoritized counterparts could not [37–39, 70, 96]. Both of those situations deserve attention, and HCI research should continue focusing study in those areas. However, we argue that *the intersection of power that politicians have on social media—and the ways that their inappropriate speech can change the world—is a special nexus that we must better contend with outside of pre-existing frameworks for studying "the powerful."* Recognizing this nexus of power also contributes to prior CSCW work on different types of online harms and how they may impact users' attitudes on moderation [47, 85] by urging for specific consideration of politicians' potential for causing harm at a grander scale than average users. Users who are not elected officials—even the most popular among them—do not hold that same type of sociopolitical power, and arguably cannot extend harm in the same way.

While participants did not have any singular takeaway on “what to do” with politician speech online—and indeed, seemed quite conflicted about it, as is outlined below—it is clear that social media users generally recognize a) the outsized influence that politicians’ inappropriate social media posts now have, and b) the ways that money and changes in the political economy of media have gotten us here.

5.1.2 The Practical and Legal Considerations of Moderating Politicians. The relationship between politicians’ freedom of speech and social media platforms, as discussed by these participants, was not only nuanced but also ever-shifting. Participants who indicated on the pre-screening survey that they “strongly agree” that platforms should be able to ban politicians’ accounts often left with trepidation on that decision, and vice versa. This makes the “findings” around this relationship unclear, but it does indicate an opportunity for revisiting, legally, the ways that platforms are understood to operate as public conduits for speech.

There was obvious dissonance in the argument over how free speech applies to politicians’ online speech, sometimes even within reflections of the same participant. The risks versus benefits of seeing politicians’ posts exactly as they were published—without any moderation for hate speech, misinformation, etc.—seemed impossible to weigh against one another for many participants, Democrat and Republican, alike. On one hand, there was belief in the public right to know what elected officials are saying, so that you can, as P19 stated, “*let your enemy tell you who he is.*” On the other hand, the detrimental effects of inappropriate politician speech were widely discussed, as well, and indicated participant support for content moderation of politicians’ accounts—just the same as regular users are subject to. Mixed into this conversation were many, often-at-odds sentiments: fears around how quickly authoritarianism can gain steam when speech suppression begins; traditional, overt belief that free speech is free speech, and should not be impeded anywhere; and a belief that politicians should not be on social media, at all.

The all-important background to this conversation is that *many people do believe social media is the new “public square,” for better or (usually) for worse.* Even most of the participants that strongly felt social media platforms should retain their current right to moderate any speech they want due to their status as “private” (publicly-traded, but non-governmental) companies shared concerns about suppressing politicians’ speech. In other words, they don’t necessarily believe that it is alright to suppress speech—they just don’t want to infringe on companies’ rights to operate as they wish. In this case, “operating as they wish” covers social media companies’ implementations of bounds of acceptable speech through their terms of agreement and community guidelines.

Taken together, *this indicates that this group of participants believe that there may be a special relationship between the First Amendment, politicians, and social media platforms that would not ordinarily exist for other types of entities.* This supports the need for the current legal struggles [12, 46] to revisit how we define the “public” nature of social networking sites by indicating that everyday social media users may already conceptualize that platforms do act as contemporary public forums—a space which is publicly-owned not in the stockholder sense, as is currently true, but in the sense of traditional public forums like streets and public parks, which are open and available for any type of constitutional discourse.

This study also makes clear the necessity of the ongoing legal scholarship that seeks to better understand how politicians’ speech, in particular, operates on these sites [8, 40]: is it “official” government business, or is it something else? And how might that impact the ways that their speech is moderated differently than other users’? These questions are further complicated by the consideration that an individual politician may have multiple accounts. For example, if such a politician desires to use certain accounts in a *personal* rather than an *official* capacity, should they be allowed to do so?

Reading between the lines, *it is clear that these participants did, indeed, understand politicians' online speech to be "official" government business, whether they like it or not.* The overwhelming sense that politicians' social media posts have real-world impact for policy, public opinion, voting decisions, and international relations—all parts of the job of being an elected official—support this claim. However, many participants indicated distaste for this concept both in principle and because of the politics of social media platforms, themselves. There was overt, repeated, and universal distrust of platforms and their willingness to be impartial conduits of speech. This is a concern shared by legal scholars [27], and it would have impact on how SNS fit into Section 230 of the CDA—if they are not impartial “distributors,” then their fit into this clause should be re-assessed. *Because of this, politicians' use of social media for official government business presented an uncomfortable relationship for these constituents—a relationship which is opaquely steered by money and algorithms in ways that the public is not privy to knowing.* For these reasons, some participants expressed a desire to remove politicians from social networking sites, entirely, and others encouraged creating new, government-operated online spaces for elected officials to have their own type of .gov “social media”—that is, an official, archived, regulated space which politicians can use to provide moment-by-moment communications with the public.

5.2 Platform Considerations

Participants had mixed beliefs on whether or not platforms had a “responsibility” to moderate politicians' inappropriate posts. Despite this, the ways that this moderation is currently done were largely deemed ineffective by those who feel politicians' speech should, in some way, be moderated. While a number of participants already use platforms' personalized moderation affordances such as controlling how much political content they would like to see or blocking/muting accounts, *many did not feel that personalized moderation tools are a suitable solution to inappropriate politician speech.* They often felt that *others* are more at risk of being adversely affected by these posts—not themselves—so personalized moderation would do nothing to combat these concerns. This can likely be attributed to third person effects on moderation preferences, as observed in prior research [47, 53, 81]. Despite this, platforms could present information nudges [48] about the utility of these personalized moderation tools in hopes of shifting public sentiment toward adopting them for curating a more user-centered approach to encountering political content online.

There are, of course, other user-enacted tools that could lead to the removal of content for *all* users, such as flagging a post or an account. However, *these tools are deemed futile and a waste of time when it comes to politicians' posts,* since participants felt their accounts would largely go unchecked by social media companies due to politicians' power, influence, and ability to bring in money. This folk theory—although abstract [19], incomplete, and possibly flawed—is grounded in participants' prior experiences and impressions, and shapes their flag usage against inappropriate politician speech. Even those participants who indicated that they had flagged politicians' posts before did not do so because they actually believed it would be removed, but rather because they felt they should “[*do something*]” (P1)—and that was the only “something” available to them. This supports prior research which indicates that users feel that the social media accounts of powerful individuals are specially protected from content moderation [10, 67].

Community notes do provide a space of interest for some participants. Particularly, those who advocate for social media users “doing their own research” when it comes to politicians' online speech—that is to say, they believe that users should take everything a politician says with skepticism, and seek out ways to verify their posts—seem intrigued by this concept. However, there are plenty of other participants who believe that most people do not actually “do their own research,” or if they do, it is from ideologically-congruent sources—so providing “notes” on politicians' inappropriate posts will not change minds.

The other platform-oriented solution that was presented by these participants was *removing problematic politician content/accounts—but using placeholder text that explains why*, such as “this post was removed due to misinformation,” or “this account was suspended due to continuous use of hate speech.” This, in some ways, could satisfy users’ desires to hold politicians accountable while curtailing the reach of harmful content. This could solve the problem of the increased promotion of labeled content, as found in research by Chipidza and Yan [14] and Zannettou [98] by making the post, itself, unshareable—but conveying public notice as to why posts were removed. However, the removal of content, of course, means that constituents are denied access to knowing the exact rhetoric—and thus values—of their politicians, which is something many participants indicated wanting to have access to. However, as was often pointed out, news sources would almost certainly pick up on these posts before their removal, and thus constituents could most likely find their specific text in news reporting, if they so chose.

As mentioned briefly above, another way this could be addressed is through *regulated, official archiving of every elected official’s social media posts*. Several participants brought up the utility of such a concept and framed it through the idea that politicians’ posts are “official” government business—meaning that they should be archived just the same as any other public communications. Ideally, this archive would be easily searchable and publicly accessible so that citizens could find any post—since-deleted or not—at any time, as evidence of officials’ public correspondence. While this does already happen in some capacity for some accounts under rules like the Presidential Records Act, it is not clear that *all* content—including content deleted either by the politician or by the platform—is archived [1]. In practice, it may also be useful to have dedicated curators for these archived accounts. Such curators could ensure future veracity and intelligibility of politicians’ online interactions by ensuring, for instance, that context is provided when they are replying to a post (i.e., archiving the post they are replying to *and* the reply, itself) or even, perhaps, including historical context to illuminate the circumstances in which the communication was occurring.

Relatedly, *a small group of participants felt that politicians should not be on social media platforms at all*, and instead should have an officially archived .gov-type platform which is meant only for politicians’ real-time communications with the public. As unrealistic as this may sound, there is evidence that this would, indeed, decrease political vitriol on social media. As Müller and Schwarz [76] found, after mainstream platforms deleted Donald Trump’s accounts in 2021, the overall toxicity of language and engagement with Trump’s rhetoric drastically decreased from his first- and second-degree followers. This would indicate that if politicians were not allowed on social media at all, and were instead given a different site to conduct their social media-type activities, the levels of online political toxicity may drastically decrease.

The path ahead for platforms and policy remains unclear. A key frustration that many participants expressed about this topic was a lack of transparency regarding whether and how platforms treat politician speech differently. Thus, regardless of what decisions platforms arrive at regarding how they would regulate politician speech, they should clarify in their policies their current approach to this problem, the reasoning behind it, and the unsettled issues that need resolution to further shape their decision-making. Our study surfaces that users, themselves, have complicated and contradictory views about what should be done to address the unique nexus where politicians, platforms, and First Amendment considerations intersect. These findings are the *start* of a conversation upon which scholars of law and platform governance can meet to better understand the normative rules and hopes for how politicians’ speech is handled online in the U.S. context. While our findings make it difficult to arrive at any one “solution” to this highly nuanced problem, stakeholders may begin weighing the relative costs and benefits of the policy- and platform-based solutions our

participants suggested with the knowledge that we are unlikely to arrive at an optimal solution without further study.

5.3 Limitations and Future Work

This study analyzes how social media users understand politicians' inappropriate speech online, but focus group studies are inherently limited by their method. We captured the opinions of a specific group of people, in a specific geographic location, at a specific moment in time. These data are not meant to be generalizable, but rather a deep dive into the multifaceted ways that these social media users are understanding the relationship between content moderation and politicians' online speech. Focus groups are also limited by social desirability bias, as participants' perceptions of the moderator, as well as of their co-participants, can impact the opinions that they are willing to share. Though the moderator followed best practices, there is little that can be done to eradicate this bias, and given the sensitive nature of political conversations, participants may have self-censored if they felt their opinions would not be well-received by the group. Further, given the small group sizes, it is possible that participants of marginalized identity groups may have been less willing to speak if they felt under-represented or alone in their opinion. All of this to say—power and social desirability matter in focus group settings, and could impact findings.

Additionally, the moment of data collection happened to be one of great political salience, as these groups met shortly after the re-election of U.S. President Donald Trump. As such, it would be interesting to capture longitudinal data on how sociopolitical contexts affect users' opinions on moderating politicians' inappropriate speech online. Relatedly, better understanding what types of contexts *do* shift users' opinions on the topic would be interesting, and a population-wide survey-type analysis could be a systematic way to uncover that information.

Because, at this moment, there is still much to be understood about how users wish moderation operated toward politicians, focus group studies could be a particularly good way for future scholars to map out and test how social media users react to or build various types of platform governance structures that they feel would best account for this complex issue. The scope of this study was limited in its exploratory nature, but more concerted work around the shape of these governance structures is warranted.

Finally, this study took a U.S.-centric approach, which brings with it all the legal frameworks, policy environments, and governing principles that that context holds. However, the question of moderating political elites is a problem the world over, and it builds in urgency as platforms have been, for instance, increasingly relinquishing some level of control in authoritarian States in order to do business with them [63]. Giving sustained attention to how users wish their politicians were moderated—or not—in countries with different sociopolitical structures and attachments to freedom of speech would be a fruitful path forward, and this study can provide a framework for continuing such work.

6 Conclusion

As platforms' affordances—and the discourses that circulate on them—continually shift our politics, the stakes of considering how politicians should be moderated will continue to be relevant. Reflecting a group of users' opinions at this moment in time, this study has outlined a number of concerns that users have on politicians' unchecked inappropriate speech on social media platforms. Whether they believe it should be removed or not, users' complete lack of faith in platforms to act in any way beside their best financial interests underlies nearly every aspect of these findings. They are concerned about how platforms' logics and opaque algorithms encourage sensational politician speech, and the power that politicians wield to turn that speech into a reality. Further, they are often deeply uncomfortable with definitively weighing freedom of speech against social harm—as

well as being nervous about the ways that content moderation of politicians' speech might impact the public's right to know about their politicians' stances. We argue that there is a special type of relationship between politicians and platforms that should be considered differently than other entities' platform-based power. It is a unique nexus which should encourage a definitive line of research slightly apart from *political speech* on platforms and slightly apart from *powerful entities' speech* on platforms. Considerations of how to approach content moderation for politicians is an increasingly urgent but under-analyzed area for researchers, policymakers, and platforms to contend with, and the potential for political chaos will only continue to ascend until we take this nuanced need seriously.

Acknowledgments

Acknowledgements will be added after blind peer review process is complete.

References

- [1] [n. d.]. Archived Social Media | Joseph R. Biden Jr. Presidential Library. <https://www.bidenlibrary.gov/research/archived-social-media>
- [2] Meysam Alizadeh, Fabrizio Gilardi, Emma Hoes, K. Jonathan Klüser, Maël Kubli, and Nahema Marchal. 2022. Content Moderation As a Political Issue: The Twitter Discourse Around Trump's Ban. *Journal of Quantitative Description: Digital Media* 2 (Oct. 2022). doi:10.51685/jqd.2022.023
- [3] Ruth E. Appel, Jennifer Pan, and Margaret E. Roberts. 2023. Partisan conflict over content moderation is more than disagreement about facts. *Science Advances* 9, 44 (Nov. 2023), eadg6799. doi:10.1126/sciadv.adg6799 Publisher: American Association for the Advancement of Science.
- [4] Shubham Atreja, Libby Hemphill, and Paul Resnick. 2023. Remove, Reduce, Inform: What Actions do People Want Social Media Platforms to Take on Potentially Misleading Content? *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 291:1–291:33. doi:10.1145/3610082
- [5] Amanda Baughan, Katherine Alejandra Cross, Elena Khasanova, and Alexis Hiniker. 2022. Shame on Who? Experimentally Reducing Shame During Political Arguments on Twitter. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 325:1–325:18. doi:10.1145/3555216
- [6] Bruce Bimber. 2014. Digital media in the Obama campaigns of 2008 and 2012: Adaptation to the personalized political communication environment. *Journal of information technology & politics* 11, 2 (2014), 130–150.
- [7] Ross Bonifacio, Jirassaya Uttarapong, Rae Jereza, and Donghee Yvette Wohn. 2025. Self-Promotion Practices and Context Collapse Management of Adult Content Creators on OnlyFans. *Proc. ACM Hum.-Comput. Interact.* 9, 1 (Jan. 2025), GROUP27:1–GROUP27:21. doi:10.1145/3701206
- [8] Samantha Briggs. 2018. The Freedom of Tweets: The Intersection of Government Use of Social Media and Public Forum Doctrine. *Columbia Journal of Law and Social Problems* 52, 1 (2018), 1–38.
- [9] Jie Cai, Aashka Patel, Azadeh Naderi, and Donghee Yvette Wohn. 2024. Content Moderation Justice and Fairness on Social Media: Comparisons Across Different Contexts and Platforms. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3613905.3650882
- [10] Robyn Caplan and Tarleton Gillespie. 2020. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. *Social Media + Society* 6, 2 (April 2020), 2056305120936636. doi:10.1177/2056305120936636 Publisher: SAGE Publications Ltd.
- [11] Pooja Casula and Richmond Y. Wong. 2025. "I see it, I scroll past it": Exploring Perceptions of Social Media Political Discourse Among Gen Z Young Adult Women In The U.S. *Proc. ACM Hum.-Comput. Interact.* 9, 2 (May 2025), CSCW109:1–CSCW109:29. doi:10.1145/3711007
- [12] Edoardo Celeste. 2021. Digital punishment: social media exclusion and the constitutionalising role of national courts. *International Review of Law, Computers & Technology* 35, 2 (May 2021), 162–184. doi:10.1080/13600869.2021.1885106
- [13] Kyle Cheney. 2025. Trump's immigration 'shock and awe' is losing in the court of law. <https://www.politico.com/news/2025/05/02/trump-immigration-court-losses-00324683>
- [14] Wallace Chipidza and Jie (Kevin) Yan. 2022. The effectiveness of flagging content belonging to prominent individuals: The case of Donald Trump on Twitter. *Journal of the Association for Information Science and Technology* 73, 11 (Nov. 2022), 1641–1658. doi:10.1002/asi.24705
- [15] Nick Clegg. 2023. Ending suspension of Trump's accounts with new guardrails to deter repeat offenses. *Meta* (Jan. 2023). <https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/>

- [16] Juliet M. Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology* 13, 1 (March 1990), 3–21. doi:10.1007/BF00988593
- [17] Kate Crawford and Tarleton Gillespie. 2016. What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society* 18, 3 (March 2016), 410–428. doi:10.1177/1461444814543163 Publisher: SAGE Publications.
- [18] W. PHILLIPS DAVISON. 1983. The Third-Person Effect in Communication. *Public Opinion Quarterly* 47, 1 (Jan. 1983), 1–15. doi:10.1086/268763
- [19] Michael Ann DeVito. 2021. Adaptive folk theorization as a path to algorithmic literacy on changing platforms. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–38.
- [20] Michael A. DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. Algorithms ruin everything:# RIPTwitter, folk theories, and resistance to algorithmic change in social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3163–3174–3163–3174.
- [21] Ángel Díaz and Laura Hecht-Felella. 2021. *Double Standards in Social Media Content Moderation*. Technical Report. Brennan Center for Justice, New York University School of Law. https://www.skeyesmedia.org/documents/bo_filemanager/Double_Standards_Content_Moderation.pdf
- [22] Mattias Ekman and Andreas Widholm. 2014. Twitter and the celebritisation of politics. *Celebrity Studies* 5, 4 (Oct. 2014), 518–520. doi:10.1080/19392397.2014.981038
- [23] Jacob Erickson and Bei Yan. 2025. Content Moderation and Hate Speech on Alternative Platforms: A Case Study of BitChute. *Proc. ACM Hum.-Comput. Interact.* 9, 2 (May 2025), CSCW052:1–CSCW052:18. doi:10.1145/3710950
- [24] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "Like" It, then I Hide It: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2371–2382–2371–2382. <http://doi.acm.org/10.1145/2858036.2858494>
- [25] K. J. Kevin Feng, Xander Koo, Lawrence Tan, Amy Bruckman, David W. McDonald, and Amy X. Zhang. 2024. Mapping the Design Space of Teachable Social Media Feed Experiences. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3613904.3642120
- [26] Sarah Fisher, Beatriz Kira, Kiran Arabaghata Basavaraj, and Jeffrey Howard. 2024. Should Politicians Be Exempt from Fact-Checking? *Journal of Online Trust and Safety* 2, 2 (Feb. 2024). doi:10.54501/jots.v2i2.170 Number: 2.
- [27] Terry Flew, Fiona Martin, and Nicolas Suzor. 2019. Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy* 10, 1 (March 2019), 33–50. doi:10.1386/jdmp.10.1.33_1
- [28] Kevin Freking. 2017. Obama makes his mark as first “social media” president. <https://www.seattletimes.com/nation-world/nation-politics/obama-makes-his-mark-as-first-social-media-president/>
- [29] Sheera Frenkel and Davey Alba. 2021. In India, Facebook grapples with an amplified version of its problems. *The New York Times* (2021). <https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html>
- [30] Zihan Gao and Jacob Thebault-Spieker. 2024. Investigating Influential Users’ Responses to Permanent Suspension on Social Media. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 79:1–79:41. doi:10.1145/3637356
- [31] Sarah Gilbert. 2023. Towards Intersectional Moderation: An Alternative Model of Moderation Built on Care and Power. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sept. 2023), 1–32. doi:10.1145/3610047
- [32] Tarleton Gillespie. 2017. Governance of and by platforms. *SAGE handbook of social media* (2017), 254–278.
- [33] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*. Yale University Press.
- [34] Thomas D. Grace, Ian Larson, and Katie Salen. 2022. Policies of Misconduct: A Content Analysis of Codes of Conduct for Online Multiplayer Games. *Proc. ACM Hum.-Comput. Interact.* 6, CHI PLAY (Oct. 2022), 250:1–250:23. doi:10.1145/3549513
- [35] Catherine Grevet, Loren G. Terveen, and Eric Gilbert. 2014. Managing political differences in social media. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1400–1408. doi:10.1145/2531602.2531676
- [36] Kirsikka Grön and Matti Nelimarkka. 2020. Party Politics, Values and the Design of Social Media Services: Implications of political elites’ values and ideologies to mitigating of political polarisation through design. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 104:1–104:29. doi:10.1145/3415175
- [37] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 466:1–466:35. doi:10.1145/3479610
- [38] Camille Harris, Amber Gayle Johnson, Sadie Palmer, Diyi Yang, and Amy Bruckman. 2023. "Honestly, I Think TikTok has a Vendetta Against Black Creators": Understanding Black Content Creator Experiences on TikTok. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 320:1–320:31. doi:10.1145/3610169

- [39] Ian Hawkins, Jessica Roden, Miriam Attal, and Haleemah Aqel. 2023. Race and gender intertwined: why intersecting identities matter for perceptions of incivility and content moderation on social media. *Journal of Communication* 73, 6 (Dec. 2023), 539–551. doi:10.1093/joc/jqad023
- [40] Kathleen McGarvey Hidy. 2019. Social Media Use and Viewpoint Discrimination: A First Amendment Judicial Tightrope Walk with Rights and Risks Hanging in the Balance. *Marquette Law Review* 102, 4 (2019), 1045–1086.
- [41] Manoel Horta Ribeiro, Shagun Jhaver, Jordi Cluet-i Martinell, Marie Reignier-Tayar, and Robert West. 2025. Deplatforming Norm-Violating Influencers on Social Media Reduces Overall Online Attention Toward Them. *Proc. ACM Hum.-Comput. Interact.* 9, 2 (May 2025), CSCW062:1–CSCW062:25. doi:10.1145/3710960
- [42] Manoel Horta Ribeiro, Shagun Jhaver, Savvas Zannettou, Jeremy Blackburn, Gianluca Stringhini, Emiliano De Cristofaro, and Robert West. 2021. Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–24. doi:10.1145/3476057
- [43] Jeff Horwitz. 2021. Facebook says its rules apply to all. Company docs reveal a secret elite that's exempt. https://www.wsj.com/tech/facebook-files-xcheck-zuckerberg-elite-rules-11631541353?mod=article_inline
- [44] Jeff Horwitz. 2021. Facebook says its rules apply to all. Company documents reveal a secret elite that's exempt. *Wall Street Journal* 13 (2021).
- [45] Y. Hua, T. Ristenpart, and M. Naaman. 2020. Towards Measuring Adversarial Twitter Interactions against Candidates in the US Midterm Elections. In *Proceedings of the International AAAI Conference on Web and Social Media*.
- [46] David L. Hudson. 2018. In the Age of Social Media, Expand the Reach of the First Amendment. *Human Rights* 43, 4 (2018), 2–4. <https://www.jstor.org/stable/27007373>
- [47] Shagun Jhaver. 2025. Bans vs. Warning Labels: Examining Bystanders' Support for Community-wide Moderation Interventions. *ACM Trans. Comput.-Hum. Interact.* 32, 2 (April 2025), 18:1–18:30. doi:10.1145/3715116
- [48] Shagun Jhaver. 2025. Personal Moderation Configurations on Facebook: Exploring the Roles of Fear of Missing Out, Social Media Addiction, Norms, and Platform Trust. *First Monday* 30, 9 (Sep. 2025). doi:10.5210/fm.v30i9.14272
- [49] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. "Did You Suspect the Post Would Be Removed?": Understanding User Reactions to Content Removals on Reddit. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 192 (Nov. 2019), 33 pages. doi:10.1145/3359294
- [50] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. *ACM Trans. Comput.-Hum. Interact.* 26, 5 (July 2019), 31:1–31:35. doi:10.1145/3338243
- [51] Shagun Jhaver, Christian Boylston, Diyi Yang, and Amy Bruckman. 2021. Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 381:1–381:30. doi:10.1145/3479525
- [52] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact.* 25, 2 (March 2018), 12:1–12:33. doi:10.1145/3185593
- [53] Shagun Jhaver, Alice Qian Zhang, Quan Ze Chen, Nikhila Natarajan, Ruotong Wang, and Amy X. Zhang. 2023. Personalizing Content Moderation on Social Media: User Perspectives on Moderation Choices, Interface Design, and Labor. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 289:1–289:33. doi:10.1145/3610080
- [54] Jialun 'Aaron' Jiang, Skyler Middler, Jed R. Brubaker, and Casey Fiesler. 2020. Characterizing Community Guidelines on Social Media Platforms. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 287–291. doi:10.1145/3406865.3418312
- [55] N. F. Johnson, R. Leahy, N. Johnson Restrepo, N. Velasquez, M. Zheng, P. Manrique, P. Devkota, and S. Wuchty. 2019. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* 573, 7773 (Sept. 2019), 261–265. doi:10.1038/s41586-019-1494-7
- [56] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic folk theories and identity: How TikTok users co-produce knowledge of identity and engage in algorithmic resistance. *Proceedings of the ACM on human-computer interaction* 5, CSCW2 (2021), 1–44.
- [57] Nayana Kirasur and Shagun Jhaver. 2025. Understanding the Prevalence of Caste: A Critical Discourse Analysis of Caste-based Marginalization on X. <https://www.shagunjhaver.com/research/articles/kirasur-2025-caste/> Section: research.
- [58] Kate Klonick. 2018. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review* (2018), 1598–1670.
- [59] Sena A. Kojah, Ben Zefeng Zhang, Carolina Are, Daniel Delmonaco, and Oliver L. Haimson. 2025. "Dialing it Back:" Shadowbanning, Invisible Digital Labor, and how Marginalized Content Creators Attempt to Mitigate the Impacts of Opaque Platform Governance. *Proc. ACM Hum.-Comput. Interact.* 9, 1 (Jan. 2025), GROUP12:1–GROUP12:22. doi:10.1145/3701191

- [60] Yubo Kou and Xinning Gui. 2021. Flag and Flaggability in Automated Moderation: The Case of Reporting Toxic Behavior in an Online Game Community. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3411764.3445279
- [61] Anastasia Kozyreva, Stefan M. Herzog, Stephan Lewandowsky, Ralph Hertwig, Philipp Lorenz-Spreen, Mark Leiser, and Jason Reifler. 2023. Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences* 120, 7 (Feb. 2023), e2210666120. doi:10.1073/pnas.2210666120
- [62] Emily Kubin, , and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45, 3 (July 2021), 188–206. doi:10.1080/23808985.2021.1976070 Publisher: Routledge.
- [63] Adrienne LaFrance. 2024. The rise of techno-authoritarianism. <https://www.theatlantic.com/magazine/archive/2024/03/facebook-meta-silicon-valley-politics/677168/>
- [64] Mireille Lalancette and Vincent Reynaud. 2019. The Power of Political Image: Justin Trudeau, Instagram, and Celebrity Politics. *American Behavioral Scientist* 63, 7 (June 2019), 888–924. doi:10.1177/0002764217744838
- [65] Wyatt Larkin. 2021. Lemmon v. Snap, Inc.: Ninth Circuit chips away at tech companies' Section 230 immunity. <https://jolt.law.harvard.edu/digest/lemmon-v-snap-inc-ninth-circuit-chips-away-at-tech-companies-section-230-immunity>
- [66] Brian D Loader, Ariadne Vromen, and Michael A Xenos. 2016. Performing for the young networked citizen? Celebrity politics, social networking and the political engagement of young people. *Media, Culture & Society* 38, 3 (April 2016), 400–419. doi:10.1177/0163443715608261
- [67] Renkai Ma and Yubo Kou. 2021. "How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 429:1–429:25. doi:10.1145/3479573
- [68] Renkai Ma, Yue You, Xinning Gui, and Yubo Kou. 2023. How Do Users Experience Moderation?: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 278:1–278:30. doi:10.1145/3610069
- [69] Samuel Mayworm, Michael Ann DeVito, Daniel Delmonaco, Hibby Thach, and Oliver L Haimson. 2024. Content moderation folk theories and perceptions of platform spirit among marginalized social media users. *ACM Transactions on Social Computing* 7, 1-4 (2024), 1–27.
- [70] Samuel Mayworm, Michael Ann DeVito, Daniel Delmonaco, Hibby Thach, and Oliver L. Haimson. 2024. Content Moderation Folk Theories and Perceptions of Platform Spirit among Marginalized Social Media Users. *Trans. Soc. Comput.* 7, 1-4 (March 2024), 1:1–1:27. doi:10.1145/3632741
- [71] Brian McInnis, Leah Ajmani, Lu Sun, Yiwen Hou, Ziwen Zeng, and Steven P. Dow. 2021. Reporting the Community Beat: Practices for Moderating Online Discussion at a News Website. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–25. doi:10.1145/3476074
- [72] Corrado Monti, Matteo Cinelli, Valensise, Walter Quattrociochi, and Michele Starnini. [n. d.]. Online conspiracy communities are more resilient to deplatforming | PNAS Nexus | Oxford Academic. <https://academic.oup.com/pnasnexus/article/2/10/pgad324/7332079>
- [73] David L. Morgan. 1993. *Successful Focus Groups: Advancing the State of the Art*. SAGE Publications, Inc. doi:10.4135/9781483349008
- [74] David L. Morgan. 1996. Focus Groups. *Annual Review of Sociology* 22 (1996), 129–152. <https://www.jstor.org/stable/2083427> Publisher: Annual Reviews.
- [75] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (Nov. 2018), 4366–4383. doi:10.1177/1461444818773059 Publisher: SAGE Publications.
- [76] Karsten Müller and Carlo Schwarz. 2022. The Effects of Online Content Moderation: Evidence from President Trump's Account Deletion. *SSRN Electronic Journal* (2022). doi:10.2139/ssrn.4296306
- [77] Matti Nelimarkka, Jean Philippe Rancy, Jennifer Grygiel, and Bryan Semaan. 2019. (Re)Design to Mitigate Political Polarization: Reflecting Habermas' ideal communication space in the United States of America and Finland. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 141:1–141:25. doi:10.1145/3359243
- [78] Christina A. Pan, Sahil Yakhmi, Tara P. Iyer, Evan Strasnick, Amy X. Zhang, and Michael S. Bernstein. 2022. Comparing the Perceived Legitimacy of Content Moderation Processes: Contractors, Algorithms, Expert Panels, and Digital Juries. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (March 2022), 1–31. doi:10.1145/3512929
- [79] Perspective. 2025. Attributes & Languages. https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US
- [80] Newley Purnell and Jeff Horwitz. 2020. Facebook's hate-speech rules collide with Indian politics. *The Wall Street Journal* (2020). <https://www.wsj.com/tech/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346>

- [81] Martin J. Riedl, Whipple, Kelsey N., and Ryan Wallace. 2022. Antecedents of support for social media content moderation and platform regulation: the role of presumed effects on self and others. *Information, Communication & Society* 25, 11 (Aug. 2022), 1632–1649. doi:10.1080/1369118X.2021.1874040 Publisher: Routledge eprint: <https://doi.org/10.1080/1369118X.2021.1874040>.
- [82] Sarah T. Roberts. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press.
- [83] Andrea Roemmele and Rachel Gibson. 2020. Scientific and subversive: The two faces of the fourth era of political campaigning. *New Media & Society* 22, 4 (April 2020), 595–610. doi:10.1177/1461444819893979 Publisher: SAGE Publications.
- [84] Morgan Klaus Scheuerman, Stacy M. Branham, and Foad Hamidi. 2018. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 155:1–155:27. doi:10.1145/3274424
- [85] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R. Brubaker. 2021. A Framework of Severity for Harmful Content Online. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 368:1–368:33. doi:10.1145/3479512
- [86] Farhana Shahid and Aditya Vashistha. 2023. Decolonizing Content Moderation: Does Uniform Global Community Standard Resemble Utopian Equality or Western Power Hegemony?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3544548.3581538
- [87] Janjira Sombatpoonsiri and Sangeeta Mahapatra. 2024. Regulation or Repression? Government Influence on Political Content Moderation in India and Thailand. *Carnegie Endowment for International Peace* (2024). <https://carnegieendowment.org/research/2024/07/india-thailand-social-media-moderation?lang=en>
- [88] Nadine Strossen. [n. d.]. A first amendment right to social media? *TalksonLaw* ([n. d.]). <https://www.talksonlaw.com/briefs/does-the-first-amendment-require-social-media-platforms-to-grant-access-to-all-users>
- [89] Amaury Trujillo and Stefano Cresci. 2022. Make Reddit Great Again: Assessing Community Effects of Moderation Interventions on r/The_Donald. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 526:1–526:28. doi:10.1145/3555639
- [90] Aleksandra Urman, Aniko Hannak, and Mykola Makhortykh. 2024. User Attitudes to Content Moderation in Web Search. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 146:1–146:27. doi:10.1145/3637423
- [91] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What It Wants": How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–22. doi:10.1145/3415238
- [92] Kristen Vaccaro, Ziang Xiao, Kevin Hamilton, and Karrie Karahalios. 2021. Contestability For Content Moderation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 318:1–318:28. doi:10.1145/3476059
- [93] Chris Vallance. 2025. Meta is ditching fact checkers for X-style community notes. Will they work? *BBC* (Jan. 2025). <https://www.bbc.com/news/articles/c4g93nvrzd7o>
- [94] Anh V. Vu, Alice Hutchings, and Ross Anderson. 2024. No Easy Way Out: the Effectiveness of Deplatforming an Extremist Forum to Suppress Hate and Harassment. doi:10.48550/arXiv.2304.07037 arXiv:2304.07037 [cs].
- [95] Leijie Wang, Kathryn Yurechko, Pranati Dani, Quan Ze Chen, and Amy X. Zhang. 2025. End User Authoring of Personalized Content Classifiers: Comparing Example Labeling, Rule Writing, and LLM Prompting. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3706598.3713691
- [96] Ina Weber, João Gonçalves, Gina M. Masullo, Marisa Torres da Silva, and Joep Hofhuis. 2024. Who Can Say What? Testing the Impact of Interpersonal Mechanisms and Gender on Fairness Evaluations of Content Moderation. *Social Media + Society* 10, 4 (2024), 20563051241286702. doi:10.1177/20563051241286702 arXiv:<https://doi.org/10.1177/20563051241286702>
- [97] Brita Ytre-Arne and Hallvard Moe. 2021. Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society* 43, 5 (2021), 807–824.
- [98] Savvas Zannettou. 2021. "I Won the Election!": An Empirical Analysis of Soft Moderation Interventions on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 15 (May 2021), 865–876. doi:10.1609/icwsm.v15i1.18110

A Participant Information

Tables 2-9 show the composition of each of our eight focus group, respectively. We show our participants' demographic details and their answers to the most relevant open-ended questions. Note that during our analysis, we assigned the participant numbers based upon the chronological ordering of group meetings.

Table 2. Disagree Group 1

	P1	P2	P3
Age	45-54	55-64	45-54
Gender	Man	Man	Woman
Race/Ethnicity	Black or African American	Black or African American	White
Political Affiliation	Independent leaning Democrat	Strong Republican	Independent leaning Republican
SNS Use/Day	4+ hours	0-2 hours	0-2 hours
Politician Ban Support?	Somewhat disagree	Strongly disagree	Somewhat disagree
Short Answer Response	The public needs to know not just the stance on where the politician stand, but also there temperament and personality. Gives a true picture for fitness for office.	Not just the 1st Amendment, but the fact that they use the public's internet requires them to post, peint, etc. every point of view. Strange, will post KKK stuff, but a Trump tweet was dangerous?	I believe it is my responsibility to determine what is harmful. I want to know both sides of issues regardless of being inappropriate. I think allowing on my right to know is bias and a controlled agenda?

Table 3. Disagree Group 2

	P4	P5	P6	P7	P8	P9
Age	18-24	65+	45-54	55-64	65+	45-54
Gender	Woman	Man	Woman	Woman	Woman	Man
Race/ Ethnicity	American Indian or Alaska Native; White	White	Black or African American	Black or African American	White	Black or African American
Political Affiliation	Independent	Independent leaning Republican	Not very strong Democrat	Independent leaning Democrat	Strong Democrat	Independent leaning Democrat
SNS Use/Day	2-4 hours	0-2 hours	4+ hours	0-2 hours	2-4 hours	4+ hours
Politician Ban Support?	Somewhat disagree	Neither agree nor disagree	Somewhat disagree	Somewhat disagree	Somewhat disagree	Somewhat disagree
Short Answer Response	The media, to a certain extent, should be a place where knowledge and opinions are shared. When the media starts to be moderated, who gets to decide what can stay and what can't? Not all media is good, but freedom of expression is vital to protect democracy.	I think they all lie but they have a right to speak	You should be free to express your opinion.	Everybody should have their own opinion.	Freedom of speech is important.	I agree they should be able to ban if they are a privately owned company or corporation.

Table 4. Disagree Group 3

	P21	P22	P23
Age	65+	45-54	35-44
Gender	Man	Man	Man
Race/Ethnicity	White	White	White
Political Affiliation	Strong Democrat	Independent leaning Democrat	Independent
SNS Use/Day	2-4 hours	4+ hours	0-2 hours
Politician Ban Support?	Disagree	Somewhat disagree	Somewhat disagree
Short Answer Response	The First Amendment	I don't think banning is the right option, though everyone does need to understand that these are private websites and thus have the right to do whatever they like. I also believe it's important to know what politicians are saying, though that can be learned elsewhere.	I think politicians speech is part of the public record and should be treated similar to other government records in terms of preservation and access.

Table 5. Disagree Group 4

	P24	P25	P26	P27
Age	18-24	35-44	45-54	45-54
Gender	Transgender Woman	Woman	Woman	Woman
Race/Ethnicity	White	Black or African American	Black or African American	White
Political Affiliation	Independent	Independent leaning Democrat	Strong Republican	Independent Leaning Republican
SNS Use/Day	4+ hours	4+ hours	2-4 hours	2-4 hours
Politician Ban Support?	Somewhat disagree	Somewhat disagree	Somewhat disagree	Somewhat disagree
Short Answer Response	It really depends on if the politician is currently in office or in the midst of running a campaign. If that's the case, I think it's wrong to ban their accounts unless they are removed from office. Otherwise, those accounts should not be removed. If they are a former politician, at the end of the day it's up to the social media platforms terms of service, which is its own can of worms, but it's the world we live in.	We should be able to stay informed about potential risks.	I don't think politician's should be banned from exercising their speech on either side. It is up to the consumer to be intelligent enough & conduct the fact finding mission necessary to come to an informed decision. This is why freedom of speech exist. The public has to be more willing to have an open mind & do the research necessary to come to an understanding & informative decision.	Some things are relative and up to interpretation

Table 6. Agree Group 1

	P10	P11	P12
Age	35-44	55-64	25-34
Gender	Man	Man	Woman
Race/Ethnicity	White	White	White
Political Affiliation	Strong Democrat	Strong Democrat	Strong Democrat
SNS Use/Day	0-2 hours	None at all	4+ hours
Politician Ban Support?	Strongly agree	Somewhat agree	Somewhat agree
Short Answer Response	A growing number of politicians are publicizing and expounding upon stances and beliefs that are seriously bigoted and biased against a variety of groups of people, this level of bigotry must be and needs to be controlled	Irresponsible, inflammatory posts likely to invoke violence or exacerbate extreme prejudice should be banned. Obviously difficult to define specifically (and perhaps, unfortunately, impossible) but political leadership in a democracy demands higher responsibilities.	If their language is used to target groups of people and is insulting rather than constructive, the language should be restricted and flagged.

Table 7. Agree Group 2

	P13	P14	P15
Age	25-34	55-64	55-64
Gender	Man	Woman	Woman
Race/Ethnicity	Asian; Native Hawaiian or Pacific Islander	Black or African American	Black or African American
Political Affiliation	Independent leaning Republican	Independent leaning Democrat	Not very strong Democrat
SNS Use/Day	0-2 hours	2-4 hours	4+ hours
Politician Ban Support?	Somewhat agree	Somewhat agree	Strongly agree
Short Answer Response	To help avoid bias and unfair influences	Politicians words and views are often incendiary, which encourages pack mentality. (1/6) insurrection	Because they should be more responsible, they pass laws and create them

Table 8. Agree Group 3

	P16	P17	P18	P19	P20
Age	65+	45-54	65+	55-64	65+
Gender	Man	Man	Man	Man	Woman
Race/Ethnicity	White	Black or African American	Black or African American	White	White
Political Affiliation	Independent leaning Democrat	Strong Democrat	Strong Democrat	Independent leaning Democrat	Strong Democrat
SNS Use/Day	2-4 hours	2-4 hours	4+ hours	2-4 hours	0-2 hours
Politician Ban Support?	Somewhat agree	Somewhat agree	Agree	Somewhat agree	Somewhat agree
Short Answer Response	I, secretly, hope that readers would be able discern when a politician's opinion has any relevance.	If politicians use false information to the public could be very problematic for everyone.	post should be non-political	I think that platforms are determiners of what they consider harmful or dangerous speech and that there need to be some limits to such speech. Obviously this needs to be balanced with freedom of expression and the right of the public to know about what candidates and representatives believe	Freedom of speech doesn't have to include hate speech or slander.

Table 9. Agree Group 4

	P28	P29	P30
Age	25-34	25-34	45-54
Gender	Woman	Woman	Woman
Race/Ethnicity	Black or African American	American Indian or Alaska Native; White	Black or African American
Political Affiliation	Independent leaning Democrat	Strong Democrat	Strong Democrat
SNS Use/Day	4+ hours	4+ hours	4+ hours
Politician Ban Support?	Strongly agree	Somewhat agree	Strongly agree
Short Answer Response	There's a thin line between freedom of speech and hate speech. Global leaders should not be able to use their platforms to promote hate. In the steps to both genocide and fascism, hate speech is always the starting point.	Everyone has a right to free speech and every person should have access to information on politicians. However, when someone engages in dangerous speech that promotes violence and hate or that intentionally causes misinformation campaigns, they should face consequences and eventual banning from platforms if they continue.	Spreading hate is wrong. If your speech could incite violence it should not be shared. We have a right to free speech but social media formats are owned by private corporations so they can ban anyone they choose.

B Focus Group Structure

As discussed in Methods, these focus groups ran on a semi-structured protocol. General flow of conversation is outlined below.

B.1 General Social Media Use

We started each focus group session with general questions about participants' social media use—their habits, platforms of choice, and approximate daily usage. We additionally asked about their propensity to follow political elites on these platforms, assessing whether they actively sought out these accounts or saw politicians' posts through algorithmically curated news feeds and/or by way of other accounts re-posting content.

B.2 Defining Inappropriate Politician Speech

A few days prior to their arrival at the focus group session, we prompted participants to make note of a politician's social media post that they found "inappropriate for any reason." During the session, we discussed these posts in concrete or abstract terms and inquired how participants would define "inappropriate" for a politician. Together, each group built a rough framework of speech patterns that they deemed inappropriate, as well as how this differed (or not) from expectations for average citizens' speech on social media.

B.3 Politicians and Content Moderation

We then moved into a discussion on participants' understanding of and opinions on content moderation of the powerful. After sharing a definition and some examples of content moderation mechanisms, participants discussed their past experience using these tools (or not) on politicians'

content, their thoughts on platform-enacted content moderation of political elites, and concerns about the bounds of the First Amendment.

B.4 Responsibility and Solutions

Finally, participants were encouraged to collaborate on thinking through new tools that might be useful for enacting content moderation of political elites, or otherwise holding them accountable for their online speech. We ended each meeting with a reflection on whose responsibility it is to hold these elites to account, and what form that responsibility might take.

Received 12 May 2025; revised 25 August 2025; accepted 11 November 2025