E-Rulemaking & Democracy: Examining Public Comment Participation from Individuals on the 2016 Broadband Privacy Order

James Seddon

Political Science Advisor: Anthony Fowler

Political Science Preceptor: Patricia Conley

Public Policy Preceptor: Rachel Dec

Joint BA Thesis, Public Policy Studies and Political Science

The University of Chicago

April 2021

Abstract

While the transition to e-rulemaking has doubtlessly increased the quantity of individual public participation in the notice and comment process, the benefits that proponents claimed the move to e-rulemaking would bring — heightened public understanding and an enhancement of the democratic process — have arguably failed to materialize. Through the use of a novel dataset constructed from the 2016 Federal Communications Commission's Broadband Privacy Order docket, Census block group level demographics data, and FCC broadband accessibility data, this paper furthers the existing literature on individual participation in public comment by examining three key questions: what sorts of individuals participate, how individuals choose to participate, and what interest group organizations motivate individuals to participate (if any). Through the use of these data and comparisons to the population by the use of a public poll and voter file data, the paper finds evidence in contrast to the existing literature that, when controlling for other demographic variables, women are no more likely to submit mass comments than men, and that non-white individuals are more likely to participate more and oppose the Order than white individuals. The paper additionally finds evidence that those with broadband are more likely to oppose the Order than those without, as well as that Republicans are more likely to participate, submit comments more frequently, and submit mass comments than non Republicans on the Order.

Acknowledgements: Academically, I'd like to thank, in no particular order, my incredible PLSC preceptor Professor Conley, my razor-sharp advisor Professor Fowler, and my very understanding PBPL preceptor Rachel Dec. Personally, I'd like to thank Emily Young for her continued support in all things over these last three and a bit very long years, my dear and longest friend Gregory Wong, the peerless Thadeus J. Obora, my oft-distracting roommates, and the endless list of other friends and peers who helped me along the way. I'd also like to thank my infinitely supportive parents, my wonderful sister, and of course, the Academy.

Table of Contents

Introduction

Established in 1934, the Federal Communications Commission (FCC) is an independent US federal regulatory agency charged with regulating interstate and international radio, television, satellite, cable, and wire communications across the United States. The FCC is led by a set of five commissioners appointed by the President and confirmed by the Senate, with one commissioner appointed by the President acting as Chairman.[1] As the primary authority on "the nation's communications law, regulation, and technological innovation," in practice the FCC makes rules that regulate broadband competition and innovation, allocates uses for most of the electromagnetic spectrum, supervises defense of the nation's communications infrastructure, and sets media regulations.[2]

Any such rulemaking actions — in the form of Commission reports and orders — require a simple majority of commissioners to enact, and are governed by the notice and comment process of the Administrative Procedures Act (APA).[3] This requires the FCC to present proposed rulemaking documents to the public before holding a Commission-wide vote to adopt the rule; in theory, it also allows anyone to submit relevant thoughts, concerns, and/or suggestions to the agency on the rule in order to provide new knowledge or to challenge the agency's conclusion on an issue. Prior to the rise of the Internet, agencies were not required to post any rulemaking documents online, nor to accept comments on the docket during the notice and comment period.[4] In practice, this paper-only method restricted rulemaking docket viewing and commenting access to only those willing to make the trek to the federal agency; unsurprisingly, these were mostly interest groups, trade organizations, and companies whom the rules would affect. Beginning in 2002, the E-Government Act required agencies to post rulemaking documents online and allow electronically submitted comments to be posted to the docket, which began a slow change towards promoting increased citizen

[1] "The Federal Communications Commission (FCC)."
[2] "What We Do."
[3] Ibid.
[4] E-Government Act of 2002.

involvement in the rulemaking process. This move towards what was branded e-rulemaking was harkened as a way of increasing participatory democracy, in contrast to the previous interest-group-led practice. According to proponents, e-rulemaking lowered the costs of participation, which would enhance the democratic process and lead to a whole host of lofty benefits: increased bureaucratic legitimacy, strengthened rights of self-governance, and a heightened public understanding of the rulemaking process. As such, policymakers strove to increase the quantity of individual comments on rulemaking proceedings.[5]

Though some issues taken up by the FCC are highly technical in nature and only affect the procedures of a niche industry, often generating few comments that mostly come from organizations, other issues that involve broadband or the media necessarily will affect almost every individual in the country and thus have a high potential for numerous individual comments. One such issue that this paper examines, the FCC's 2016 *Protecting the Privacy of Customers of Broadband and Other Telecommunications Services Report and Order* (or Broadband Privacy Order), would have implemented strong consumer data privacy regulations to all companies providing broadband service in the United States. In sum, the Order would have required consumer broadband carriers to provide key details to consumers regarding what information about them was collected and shared with third parties like advertisers, as well as mandating an opt-in process for broadband companies to share or use parts of that data. As 92.7% of all Americans had access to wired broadband in their residence in 2016 (not to mention cellular broadband), it was no surprise that this highly impactful rule would generate a docket of over 300,000 comments submitted by real individuals.[6] Yet agencies like the FCC have struggled with the implications of lowered costs of participation. Dockets such as the Broadband Privacy Order with a high number of comments often contain a high proportion of mass comments from click-through petitions or form letters as, with time, the ease of access to dockets has been co-opted for strategic use by interest groups.[7] According

[5] de Figueiredo, "E-Rulemaking," 3-8.
[6] "FCC Fixed Broadband Deployment."
[7] Ibid, 21.

to some scholars like Shulman (2009), rather than lower participation costs generating increased democratic engagement, these interest group-motivated mass comments have simply "had the effect of piling insubstantial comments containing very little new information on the desks of agency officials."[8]

With this in mind, this paper will broadly examine whether, twenty years on, e-rulemaking is accomplishing its goals of inclusivity and substantive democratic engagement through an in depth analysis of individual comments and commenters on the Broadband Privacy Order docket. Although the rulemaking process is not majoritarian in nature, the kinds of individuals who participate in rulemaking have clear and decisive implications for whose opinions and suggestions are heard by agencies like the FCC, as well as determining where any potential benefits of e-rulemaking are felt. This paper will also investigate whether, through the form and substance of public comments, individuals actually understand the goals of the rulemaking process. A significant number of values-based, rather than policy-substantive, mass comments can point to a mass belief among commenters that policymakers can or do take into account the volume of comments taking a particular stance on the issue when making policy decisions, when in fact this is legally prohibited. Both analyses are required to evaluate whether the initial goals of e-rulemaking have been met, though they also contain implications for the democratic process in their own rights.

Finally, this paper will attempt to study the sources of mass comments, with a special focus on searching for "manufactured salience," a term recently coined by Jordan and Watson (2019) to characterize a type of interest group-led mass commenting known as astroturfing. Here, individuals are prompted to express their identity through words and mechanisms recommended by interest groups, who are attempting to obfuscate the original sponsors of a message from both the individual commenters and policymakers. Manufactured salience is intended to make it seem as though a message is generated from a grassroots groundswell, and

[8] Shulman, "The Case Against Mass E-Mails," 24.

therefore to convince policymakers that an issue position is more salient than it is in reality.[9] While studying manufactured salience in this paper helps satisfy the call from Jordan and Watson (2019) for additional study on the topic, it also sheds light on the potential for manipulation of individuals and policymakers both on the Order and in general.

Ultimately, this paper provides further evidence that, twenty years on, e-rulemaking is clearly not accomplishing its goals of inclusivity and substantive democratic engagement by investigating who participates, how they participate (by original message or through mass comment), what organizations, if any, push them to participate, and how these differ by support or opposition of the rule. To do so, this paper makes use of a novel dataset constructed from individual Broadband Privacy Order docket comments — the FCC's 4th most frequently commented-on rulemaking proceeding — as well as Census block group level demographics data and FCC broadband data. These data will be used in conjunction with individual voter file data from North Carolina and Florida, as well as a nationally conducted public poll on the issue, to tease out differences between individual commenters, registered voters, and the general public.

In this paper, I first present the history and background of the Broadband Privacy Order policy, as well as explain the process of Administrative Procedures Act notice and comment rulemaking in general. I then discuss the current literature on individual participation, mass comments, and manufactured salience. I explain the methodology used in the production of the novel Broadband Privacy Order dataset and consider limitations of the research associated with ecological fallacy concerns. I analyze the dataset by comparison to the nation and a public poll, as well as find potential correlations between demographics and sentiment, frequency, and mass comment likelihood. I also present qualitative research of the source of mass comments in the dataset, and use the voter file to compare matched individual commenters with registered voters. I conclude with a discussion of findings, implications, and a related ethical concern with the dataset.

[9] Jordan and Watson, "Reexamining Rulemaking in an Era of Internet-Enabled Participation," 22.

Policy Background

Proposed Content in the Broadband Privacy Order

Adopted on October 27, 2016 by a vote of 3 Democratic-leaning commissioners in favor to 2 Republicans opposed, the FCC's *Protecting the Privacy of Customers of Broadband and Other Telecommunications Services Report and Order* (or Broadband Privacy Order) applied and re-defined the strict privacy requirements applicable to Title II common carriers in the Telecommunications Act of 1996 ("Act") to broadband service specifically.[10] Although initially passed to protect consumer privacy with telephone providers, Title II was accepted to apply to internet service providers in the early days of the internet, when internet connectivity was provided over copper phone lines. Subsequent laws failed to classify broadband internet service providers as anything (due to their novelty at the time), leading to a wild west and an eventual classification as falling under Title I in 2002 until it was changed to Title II by the FCC in 2015 in their Open Internet Order.

In sum, the FCC re-defined the information protected under Section 222 of the Act granting privacy requirements as "customer proprietary information" (customer PI); this was to be comprised of "individually identifiable customer proprietary network information" (CPNI), personally identifiable information (PII), and the content of communications.[11] To protect the privacy of customer PI, consumer broadband carriers would be required to provide privacy notices that "clearly and accurately inform[ed] customers about confidential information" collected, the categories of customer PI collected, how it was used, and finally how it was shared. Carriers would also be required to give opt-in approval to share or use customer PI, and allow opt-out approval for use of sharing non-sensitive PI. Finally, the Order also mandated specific data breach notification requirements.[12]

[10] "Broadband Privacy Order," 3.
[11] Ibid.
[12] Ibid, 4-6.

To do so, the FCC first defined customers of broadband providers as "current or former subscribers, or applicants."[13] It then re-defined the scope of customer PI. CPNI would include a customer's broadband plan, location, individual device identifier, connection equipment (e.g. a router or modem), as well as broad traffic statistics, among other more-technical components.[14] PII would be classified as "any information . . . reasonably linkable to an individual or device," which would include but not be limited to:

> "Social Security number; date of birth; mother's maiden name; government-issued identifiers . . .; physical [location]; . . . online contact information; phone numbers; . . . unique device identifiers; IP addresses; and persistent online or unique advertising identifiers."[15]

Further, the FCC defined customer PI to include the content of communications, as it found that "content [was] highly individualistic, private, and sensitive."[16] This itself was sweepingly defined to mean "any part of the substance . . . or meaning of a communication or part of a communication."[17] Finally, the FCC defined sensitive customer PI as the content of communications (ex: browsing history and app usage history) and (broadly) PII.[18]

This push for privacy reform in broadband was motivated by two important factors. First, the three Democratic-leaning Commissioners voting for the measure mentioned in varying forms their belief that "it's your data. How it's used and shared should be your choice."[19] Commissioner Clyburn specifically pointed out that, in spite of this, "ninety-one percent of Americans believe, [sic] consumers have lost control of how their personal information is collected, [sic] and used by companies."[20] Resolving this concern by regulating consumer broadband providers would help because, unlike "edge carriers" like Google, broadband

[13] Ibid, 16.
[14] Ibid, 21.
[15] Ibid, 34-36.
[16] Ibid, 39.
[17] Ibid, 40.
[18] Ibid, 70-78.
[19] Ibid, 204.
[20] Ibid, 205.

providers could hypothetically "see 100% of unencrypted network traffic." While encryption of HTTP requests could help protect consumer privacy, broadband providers still could access lots of consumer information that edge carriers could not, like which websites visited were and for how long.[21] Additionally, unlike with browser extensions that could control third-party tracking, without the benefit of "robust competition in the fixed broadband market" that might allow for a switch to a new provider that did not track, most Americans had no option to opt-out.[22] Finally, the collection and use of this private information was not just a hypothetical; the Commission points out two specific examples — AT&T and Verizon — that had advertised service where, in order to receive a lower price, consumers were required to opt into sharing individual browsing data (including search terms and web pages visited) so they could tailor ads.[23] Clearly, then, Americans wanted more of a say in how their data was collected and used, and this Order would have helped resolve this problem. Yet in addition to the desire for increased privacy, thanks to the earlier reclassification of broadband providers to Title II carriers, Section 222 mandated a legal duty to protect customers' PI. Thus, since the Commission now had a statutory responsibility to oversee privacy practices — ceded from the FTC thanks to the Title II move — it was also required to act in some manner to clarify carrier responsibilities.[24]

Process and Purpose of Notice and Comment Rulemaking

The Administrative Procedure Act outlines the process of formulating, amending, issuing, or repealing regulations for federal agencies, and serves as the backbone of the administrative state.[25] All agencies, including the Federal Communications Commission, (usually) go through a standardized three part process upon beginning the exercise of issuing regula-

[21] Ibid, 11.
[22] Ibid, 13.
[23] Ibid, 77.
[24] Ibid, 8.
[25] Gangadharan, "Public Participation and Agency Discretion in Rulemaking at the Federal Communications Commission," 4.

tions. First, the government publishes documentation for proposed regulations in the form of Notices of Proposed Rulemaking (NPRMs). These NPRMs are required to specify the legal authority an agency has to implement a rule, the duration of the comment period, and information about the proposed rule the agency wants to implement. Upon the issuance of this NPRM, any interested parties — often members of the public, companies, or industries — file public comments on the proceeding, giving their own answers to questions posed in the NPRM and their opinion of parts of the proposed regulation.[26] In the case of the Federal Communication Commission, this most commonly happens via the use of the Electronic Comment Filing System (ECFS). The ECFS serves as the "repository" for official records from the FCC and the public on any docketed proceeding. Commenters have the option of either filing using the Express Comment system, for "quick submissions by individuals," or the more complicated Standard Filing system, used to submit "more detailed submissions . . . including [those by] law firms" and frequently organizations. Commenters also have the right to send paper documents into the FCC to be included in the public comment period, although these are scanned into the ECFS by Commission staff.[27] The Express Comment system, like other online e-rulemaking systems (regulations.gov, etc) was designed with the intention of moving away from a paper notice-and-comment-based rulemaking system dominated by regulated entities and associated interest groups and towards increased participation from the broader public.[28] Unsurprisingly, then, the Express Comment system — allowing for submission of a comment with about the same ease as a traditional forum post — is most frequently used by individuals; it requires the submission of the filer's name, physical address, and comment, as well as optionally an email address.[29]

Public comments submitted to the FCC are intended to both "provide the FCC with information that will increase its knowledge of the subject matter of the proposed rule,"

[26] Ibid.
[27] "ECFS FAQ."
[28] Mendelson, "Rulemaking, Democracy, and Torrents of E-Mail Annual Review of Administrative Law," 16.
[29] Ibid.

as well as "permit the public to challenge the factual assumptions, analyses, and tentative conclusions underlying the proposed rule and to show the agency the respects in which it might be in error."[30] However, any individual or group may comment on the proposed rule through the submission of data, views, or arguments; this often results in a very wide-scoped set of comments.[31] Upon completion of the public comment period, which can range from thirty to sixty days, the Commission must review responses but is not required to adhere to public opinion. Indeed, the Commission and other rulemaking agencies are actually prohibited from basing the final rule on the mere number of comments in support or opposition of the final rule, and must instead consider the merit of each unique comment as it elicits new information or raises previously-unconsidered issues.[32]

Unfortunately, this distinction of merit versus "plebiscite-style participation" is often — intentionally or not — disguised from individuals, especially when interest groups are involved in mobilization for public comment. Click-to-send, pre-filled form comments are not akin to votes; as confirmed by Shulman (2009) after discussions with more than 200 federal policymakers, "the volume of comments alone carries little or no actual weight in [actually] making a regulatory determination," let alone the textual prohibition on doing so. Yet this does not mean that individuals need not participate, even if they do not have a technical clarification to share — as discussed by Carlitz (2002) and others, greater public participation means including more points of view in the discussion, which can lead to better, more uniform, and less judicially-challengeable rules.[33] This necessarily requires a diverse array of key stakeholders — in this NPRM, which affects anyone who would at some point surf the web, the entire socioeconomic spectrum — to participate.

After considering the merit of public comments, agencies like the FCC subsequently can issue a Final Rule — called a Report and Order at the FCC — and is only required

[30] "Public Notice and Comment Rulemaking (United States)," 3.
[31] Garvey, "A Brief Overview of Rulemaking and Judicial Review."
[32] Shulman, 12.
[33] Carlitz and Gunn, "Online Rulemaking," 6.

to summarize and give reasons for dismissing "significant" arguments presented in public comments.[34] These Final Rules then have to clear White House review at OIRA before being published in the Federal Register, which importantly, is when they officially take effect as regulations. Like any piece of legislation, though, regulations can specify delayed implementation or compliance dates.[35] So-called "midnight rules," which are regulations that are finalized by federal agencies and published in the period between the election and the inauguration of a new President in order to cement the regulation and avoid potential meddling by the new administration, usually speed up these often-slow steps dramatically.[36]

Repeal and Polling-Measured Popularity of the Broadband Privacy Order

In the case of the FCC's Broadband Privacy Order, the process of issuing the regulation played out like any other. On March 31, 2016, the FCC adopted and issued an NPRM outlining its policy goals and posing questions on the best ways to implement specific provisions to any interested parties. The public comment deadline was set for May 27, 2016, and the Reply Comment — essentially rebuttals on comments — deadline was set for June 27.[37] The public comment period proved *extremely* popular — becoming the 4th most commented-on NPRM with 275,000 responses — and so the reply comment deadline was extended by 10 days to July 6.[38] Following this, the Commission took two months to read through public comments on the NPRM and refine its Report and Order; the Report and Order was adopted on October 27, 2016 — 13 days before the 2016 election. From there, things got more complicated upon the election of President Donald Trump and subsequent Republican control of both Congress and the Presidency.

[34] Garvey.
[35] O'Connell, "Agency Rulemaking and Political Transitions," 7.
[36] Ibid, 3.
[37] "FCC Releases Proposed Rules to Protect Broadband Consumer Privacy."
[38] "Broadband Privacy Order," 3.

Pursuant to the rulemaking process, by the time the Broadband Privacy Order received OIRA approval and was submitted to the Federal Register on December 2, 2016, it was clear that the incoming President and Congress would oppose the rule.[39] Yet, in opposition to typical midnight rules, though it had been officially approved between the election of a new President and their inauguration, the FCC's Broadband Privacy Order had already been passed by the FCC. Regardless, as was their legal right, the incoming Republican 115th Congress did not like the rule, claiming that the new rules "subjected [different] parts of the internet ecosystem to different rules and different jurisdictions," and that the FCC Broadband Privacy Order needed to be repealed to "restore a level playing field."[40] They subsequently began the previously-little-used process for an expedited legislative repeal of the rule known as the Congressional Review Act (CRA) upon the start of a new Congress. As the name suggests, this Act requires all federal agencies to submit a copy of any new rule to Congress, which is allowed to subsequently review it and have sixty legislative days to repeal it through an expedited process if desired.[41]

Before the 115th Congress, the CRA had only successfully been used once to repeal a 2001 Department of Labor rule, but the 115th Congress used the procedure thirteen times in the first 4 months of 2017; since the Order was submitted to Congress on December 2, 2016 (at the same time as the Federal Register), this could and did include on the Broadband Privacy Order from the FCC.[42] The joint resolution was introduced in the Senate on March 7, 2017, where it was referred to and subsequently discharged by the Senate Committee on Commerce, Science, and Transportation the following week. On the 23rd, it passed the Senate 50-48 on a party-line vote, and was referred to the House. Five days later, it passed the House 215-205, with 215 Republicans for and 190 Democrats plus 15 Republicans against.[43] The bill was then sent to and signed by the President, and the Broadband Privacy

[39] "Protecting the Privacy of Customers of Broadband and Other Telecommunications Services."
[40] House Speaker Paul Ryan, Providing For Consideration Of S.J. Res. 34, Providing For Congressional Disapproval Of A Rule Submitted By The Federal Communications Commission.
[41] Larkin, "Reawakening the Congressional Review Act," 15.
[42] U. S. Government Accountability Office, "Congressional Review Act."
[43] "Actions - S.J.Res.34 - 115th Congress (2017-2018)."

Order was subsequently repealed effective April 3, 2017.[44] Pursuant to the CRA, the FCC was also prohibited from enacting substantially similar rules in the future.

In late March of 2017 — between when the House passed the bill and when the President signed it — YouGov and the Huffington Post conducted a poll of 1000 Americans over 18 on the issue, finding that approximately 83 percent of Americans felt that telecom and cable companies should not be allowed to share personal information about customers, including browsing history, without first getting customer permission. These proportions were robust across all party id, regional, family income, and age groups, although slightly less consistently across race due to higher "not sure" responses. When asked about the FCC rules themselves, respondents felt mostly the same, with 71 percent of people agreeing that the ban should go into effect rather than be overturned and 12 percent suggesting it should be overturned; even when prompted that a Republican Congress was overturning it, this finding was robust to all aforementioned groups. Finally, when YouGov asked whether President Trump should veto the bill, 74 percent of people suggested he should, with only 11 of people suggesting he sign it. Again, this figure was robust across all groups previously mentioned.[45]

With only 11 to 12 percent of people nationally agreeing that the Broadband Privacy Order should be overturned by Congress — with only 14 to 15 percent of Republicans — it should be expected that similarly-bipartisan levels of support would be recorded during the NPRM public comment period. However, surprisingly, this was not the case; 94.5% of all comments were against the implementation of the rule, as this paper finds later in the data analysis section. This fact was not acknowledged by the FCC, which wrongly stated that "the vast majority [of FCC commenters] show support for the adoption of strong privacy rules" in its final rule.[46] Through a comparison between commenters and the population by demographics, this paper aims to examine the intriguing gap between population level sentiment and those of commenters.

[44] Ibid.
[45] "Huffington Post Online Privacy Poll."
[46] Broadband Privacy Order, 205.

Literature Review

Individual Participation in Rulemaking

While the intent of e-rulemaking was to enhance citizen participation in the notice and comment process, not all individuals are equally represented in different facets of public comments. While the particulars of why certain groups are under or overrepresented are beyond the scope of this paper, these reasons can change the demographic composition of individuals participating in the public comment. First and most obviously, differences caused by the pre-existing digital divide — a term used to describe the disparity between individuals with differing levels of Internet access, speed, and use — demonstrates a disparity between the wealthier and poorer, as well as between white individuals and other racial groups in general online participation (Remaley, 2020). Interestingly though, in a study of individual public comments on the 2016 FCC net neutrality docket, Singel (2018) finds no significant difference in participation or sentiment between those with or without broadband.[47] However, in contrast to the predominant literature on public comment that suggests including mass comments in analysis — as each contains true individual beliefs and will be read by policymakers — Singel assumes that mass commenters should be excluded from analysis due to the sheer proportional volume of their near-identical comments; this is discussed in more detail in the section below. This paper will aim to compare these findings using a new methodology and dataset, and without the potentially-flawed Singel-led complete exclusion of mass commenters.

Additionally, in a study of in-person community meetings run by the EPA on climate change in the mid-Atlantic region by Olsen, Galloway, and Ruth (2009) finds that wealthier and older individuals were proportionally much more likely to participate in in-person public comment opportunities than would be expected when compared to the general population. Interestingly, they find no proportional sex gap (relative to the population), but that the

[47] The methodology section below will discuss these concerns further.

18-44 age group was significantly underrepresented in meetings. While on face this may seem inapplicable to online rulemaking due to the differences in resources required to participate, Schlozman, Verba, and Brady (2010), examining 2008 Pew survey data, find no evidence that the relationship between political participation and socioeconomic status changes due to participation being on the internet.

A more focused study on public comments by Farina et al (2011) confirms that, despite digital divide and adoption concerns, older individuals are twice as likely to participate in an online public comment process than would be expected; similarly, those aged 49 or under were half as likely to participate as would be expected from the general population. Additionally, in contrast to the previous Olsen et al findings, a study by Bryer (2013) of the most commented-on agency notice and comment proceedings found that women were more likely than men to submit public comments by around 20 percentage points. This paper aims to test the findings of overrepresented demographics from Olsen (2009) against those of Bryer (2013) on sex, as well as to confirm the findings of Farina (2011) and Olsen on a new census block group level dataset and issue area.

Of the individuals that do submit public comments, there are clear differences between those submitting the two possible kinds of public comments: mass comments — such as form letters — and unique, originally-written comments. From a survey of 1600 individuals who had contributed to a notice and comment opportunity in the United States, Schlosberg, Zavestoski, and Shulman (2007) find a significant difference between these two groups. Those most likely to submit an originally written comment were more likely to have a higher level of education, income, and much more likely to be men; conversely, those submitting mass comments were more likely to be women, have less education, and have lower income. Interestingly, original commenters were much more likely to report a positive view of the agency in question, as well more likely to report higher satisfaction with agency decisions they commented on. Those submitting mass comments were more likely to be negative about the government in general, as well as state that their participation in public comment led

to a negative view of the agency. The authors hypothesize this lack of trust in government may explain why commenters who submit form letters are much more likely to submit a comment more than ten times as compared to original commenters. This paper will attempt to reproduce these comparative differences in sex and income by examining mass comment proportions at the census block group level.

On Mass Comments

The inclusion of mass comments in the data — which comprised 98% of all comments — raises the question of how these form letter comments, submitted en-masse, are treated by policymakers. In contrast to the APA vision, where public comments provide new knowledge to an agency, as well as allow the public to challenge the analyses and conclusions of a proposed rule, submitting form letters is an "isolated, one way, and largely unthinking" act (Schlosberg & Dryzek, 2002). Form letters are also the most common kind of comment submitted in notice and comment procedures. A study by Shulman (2009) finds that between 70% and 80% of studied public comments submitted to the EPA were completely unmodified form letters sent by members of various environmental advocacy groups; studying a particularly egregious example, Cuellar (2004) finds that, in the case of the Campaign Finance Regulatory Rule, 98.4% of all comments were from unmodified form letters. More recently, an examination by Singel (2018) of the FCC comments surrounding the Net Neutrality repeal proceedings finds that 96.5% of all comments were mass comments. For this reason, the author subsequently advocates for completely excluding all form comments from sentiment analysis, finding that, when done, over 98% of unique comments support Net Neutrality. As this paper examines the sentiment of each comment in order to describe what groups of individuals most support the Broadband Privacy Order, this is a meaningful methodological question as to whether these comments matter and should be included.

The question of whether mass comments should be included is both a question of how

policymakers react to form letters sent en-masse, as well as whether individuals submitting form letters actually agree with them in the first place. First, on policymaker reaction, Shulman (2009) reminds that, since the relevant legal statutes make clear that the "simple accumulation of mass sentiment . . . should not be a factor in making a decision," federal policymakers cannot take into account the number of mass comments at all.[48] Indeed, after surveying 200 federal agency personnel, he finds that policymakers report that the volume of comments alone carries little or no actual weight in making a decision; instead, it is the merit of the comment itself and its "unique qualities" that are valued. Since these mass comments were neither substantive nor original, Shulman reports that EPA staff were left with hundreds of thousands of comments with no new information, contributing to EPA staff "wondering how much people truly cared."[49]

Similarly, Mendelson (2011) finds that the mass comments are "sometimes derided by agency staff," who pass over comments lightly and save detailed responses (in final rules) for only the sophisticated or technical comments; rulemaking documents only occasionally acknowledge the sentiments of layperson comments at all, rarely giving them much weight. Yet, in both this piece and her 2012 follow-up, Mendelson points out that the exclusion of public comments altogether is contrary to the goal of providing additional information and arguments provided by notice and comment, even though the number of positive or negative comments alone does not affect an agency's judgement. Indeed, Mendelson points to a set of examples where various agencies have clearly reacted to mass messages, either by recognizing it indicates misunderstandings in the public consciousness and attempting to correct them, or by using it as a gauge for public resistance. As such, although policymakers do not often appreciate mass comments, it is clear they can and do use them to gauge broad public counterarguments and potential backlash if nothing else.

Next, the question of whether individuals submitting comments believe them is also

[48] Shulman, "The Case Against Mass E-Mails," 12.
[49] Ibid, 24.

nuanced. As Mendelson (2011) mentions, agency officials may reasonably be concerned with mass comments if they believe that the use of identical, or near-identical, text in a mass comment indicates it is not actually the genuinely held belief of the commenter; if nothing else, it "undermines the impression that an individual is communicating her deeply held convictions." However, this does not make the comment "content-free," and since most comments are attributable to a real person, Mendelson claims that policymakers can assume the individual is "likely to press send only if the suggested text reasonably corresponds" to their views.[50] Mendelson hypothesizes that this may be an effort to lower participation costs — and thus increase public engagement — from the interest group that drafted the form letter. As a result, so long as form messages are not from bots, this paper will assume that individuals sending messages do in fact believe in what they are sending. Since both the policymaker and individual belief conditions are satisfied, this paper will reject the Singel (2018) method and instead include mass comments — not just unique ones — in its sentiment analysis.

While the above argument clearly allows for mass comments to be included in sentiment analysis, it is also pivotally important to point out the subsequently incurred limitation that can be ascribed to the dataset. When expected issue salience with the general public is low — as it is in this paper's discussed FCC Internet Privacy rule — and yet a large number of comments submitted by individuals indicates salience is unexpectedly high, we must consider whether this is simply manufactured salience. Coined by Jordan and Watson (2019), manufactured salience occurs when "politically or economically motivated actors" generate mass comments *through online astroturfing* — the obfuscation of the original sponsors of a message in order to make it seem as though it is generated from a grassroots groundswell — to make it seem to policymakers as though an issue position is more salient than it actually is. While it is tempting to thus write this off as completely manufactured, Jordan and Watson point out that this still requires a manufacturing of salience at the individual

[50] Mendelson, "Rulemaking, Democracy, and Torrents of E-Mail," 34.

commenter level; individuals are expressing their "identities, stories, and interests" through "mechanisms and words recommended to them by external actors."[51] Thus, while a detection of highly manufactured salience in a comment dataset means it is subsequently impossible to determine true issue salience on the issue from data alone, it importantly does not mean the opinions of the commenters are somehow invalid, as also established above. Since the identities of commenters are being used to manufacture salience, this paper, which does have highly manufactured salience in its comment dataset, can and will investigate correlated demographics of mass letters with apparently-manufactured salience; however, it will not be able to draw conclusions about issue salience in the general population.

Methodology

Geocoding and Census Block Group Level Data

The main data source for this paper comes from the FCC's initial Broadband Privacy Order NPRM ECFS public comment data. There were approximately 275,000 individual responses to the proposed Order in the open April to July period of 2017, with each containing an address field, a public comment submitter name field, and a field for the comment itself. To validate and individualize these comments, I first used the FCC's API to download all comments for the NPRM within the requisite dates. This was then converted to CSV, and then merged into one large file for further analysis. The file was filtered to only include instances with actual address data — which mostly removed spam filings — leaving 246,000 unique addresses in the data.

These addresses were then cleaned using the USAddress package in Python, spliced into their component address parts, and then spliced back together, which was used to normalize the addresses and extract as much data as possible out of the often very poorly entered

[51] Jordan and Watson, "Reexamining Rulemaking in an Era of Internet-Enabled Participation," 22.

address string.[52] While highly effective, this step likely did not extract and normalize address information perfectly, resulting in some data dropping out of the dataset; however, so long as we assume that lack of clear address data between components is distributed randomly through the population, this should not be a problem. These addresses were then fed into the Census' geocoding API to validate and geocode address data. This was subsequently filtered for only unique, valid address data. After cleaning and filtering, the dataset contains 162,000 unique respondents (by address), each containing name and public comment data. This 58% yield on the starting data is likely due to the good-but-not-great Census geocoding data, as well as potential incorrect spellings of addresses by individual respondents. It is likely possible to fix a subset of these, but at 100,000 invalid entries, such a task is daunting to do manually and, considering I have exhausted existing free address cleaning and validating options, is nigh-impossible to do automatically. As such, I proceeded on as-is.

To map high-resolution demographics data onto each respondent, I then downloaded the 2010 Census Block Group-level demographic data sourced from an existing ESRI dataset, loaded it into QGIS, and proceeded to spatially match the geocoded addresses in point form from above onto it.[53] While Census Block-level data was theoretically available, I was unable to successfully load it into QGIS without crashing due to memory limitations. From the Census demographics dataset, I chose the proportions of white and male individuals living in each census block group, which were chosen as results of the literature review. I also downloaded Census Block Group-level data on broadband availability from the FCC, containing information at the Block Group level about whether broadband was available, as well as median household income data (which I converted to thousands of dollars) from the National Historical Geographic Information System.[54] These were appended by simply matching to the spatially-joined Block Group numbers in R.

[52] *Datamade/UsAddress.*
[53] "USA Census Block Group Boundaries, 2010."
[54] Manson, Steven et al., "National Historical Geographic Information System."

Precinct-Level Trump Vote Share Data

As the aim of this study was to work with individual-level data (or as closely to individual-level data as possible), I endeavored to collect the highest resolution data on party vote-share in 2016 — the closest election to the comment period — possible. Unfortunately, as I quickly discovered, there is no existing free and public dataset of a national precinct level shapefile for the 2016 election, which meant I needed to create one. I began with manually cleaning an existing shapefile dataset from the Voting and Election Science Team at the University of Florida, which eventually allowed me to match 78% of all precincts; this was a shapefile of essentially every precinct in 41 of the 50 states, with 8 states missing or unmatchable from the dataset, and one state, Georgia, partially matched.[55] I then fixed the errors in the state of Georgia from the University of Florida dataset with the MIT Election Data and Science Lab data — which contained a dataset of precinct names and results, but no shapefiles — by cleaning and matching the data in a new format.[56] This added another 1400 precincts to the dataset, and mostly completed the Georgia precincts dataset, although some rural areas were still left out of the final precinct map. I similarly matched all precinct data from Missouri onto the Voting and Election Science Team shapefiles by correcting formatting errors left by the Voting and Election Science team that had broken the shapefile, regaining most of the state precinct data in the process.

This left 7 states with no shapefiles included in the original University of Florida dataset: Connecticut, Alabama, Indiana, Mississippi, Missouri, New York, Ohio, and West Virginia. Since I only had geographic point data for each address, these shapefiles were necessary to match addresses to election precincts. Thus, I began by sourcing 2016 precinct shapefiles, and subsequently matching on the MIT Election data when available; precinct names were matched to listed precinct names in the MIT Election dataset. Unfortunately, this immediately excluded Connecticut and Alabama, whose precinct names were not available below the

[55] Voting and Election Science Team, "2016 Precinct-Level Election Results."
[56] MIT Election Data and Science Lab, "U.S. President Precinct-Level Returns 2016."

township level and in very different form respectively. I obtained a West Virginia precinct shapefile from Princeton's database, which only partially matched onto the MIT data.[57] I obtained full Indiana and Ohio precinct level shapefile and election data from the MGGG Redistricting Lab on Github.[58] This left me with Missouri and New York; using a crowd-sourced precinct shapefile collection called Election Geodata, I partially matched precinct data from Mississippi on the MIT data.

Finally, and problematically, I could not find an existing free precinct shapefile for New York. Since it contained over 7,000 responses — about 5% of the total data — and its exclusion would remove the second largest Democratic state in the country, potentially biasing conclusions derived from the FCC data, I was forced to use an available 2016 dataset containing election results and a shapefile on New York State primary results.[59] Although I spent a sizable portion of time attempting to match the precinct names in the shapefile to the MIT general election data, the names in the NY primary shapefile were completely different and ultimately unmatchable given the existing data I had. In the name of limited compatibility, I summed up all votes tallied for Republican and Democratic candidates in each primary election, and subsequently derived the ratio between each for each precinct. I am acutely aware that the use of the ratio of the sum total of primary results by party per precinct is not truly comparable to the straight choice between Clinton and Trump in the 2016 election for a multitude of possible reasons, but given the budget of this paper, the use of this ratio was the best — and only — available option besides dropping the state from analysis entirely. Future analyses using 2016 precinct shapefiles should endeavor to either construct their own New York shapefiles, or match based on the upcoming release of the 2020 Census Voting District dataset.

After collecting precinct level data on 48 of 50 states, I once again spatially matched these data against the FCC address point data, which yielded a matched result of 95.6%. While

[57] "West Virginia Voting Districts from 2010 Census."
[58] *Mggg-States/IN-Shapefiles*; *Mggg/Ohio-Precincts*.
[59] rarohla, "2016 New York Primaries by Precinct."

this was mostly due to the exclusion of Alabama and Connecticut, some of the unmatched address data were from mostly-rural areas without valid shapefile matches in the states listed above. I subsequently dropped the unmatched data; as a result, it is possible that this paper will have a slight unavoidable bias towards including more-urban areas in the public comment set. From these data, I calculated the Trump vote share relative to the Trump-Clinton total vote in each precinct, with the exception of New York (see above). While a few precincts in a few states had meaningful third-party vote share relative to the baseline, this two-party share was calculated for inter-state compatibility, as well as to exclude biases for third-party candidates only on the ballot in some states. While this measure of precinct-level party lean will not be perfectly representative of the true ideology of a district due to the exclusion of third party candidates, it should be sufficient to operationalize most of the party preference of a precinct.

Text Analysis: Mass Comment Detection and Sentiment Coding

Upon setting out to clean and code the text, I realized quickly that almost all of the comments in the dataset were non-unique form comments, although there were a handful of different form comments as discussed in the data analysis section below. In order to code these as form comments, I somewhat mimicked the spirit of the procedure laid out by Kao (2017) in his Github repository and post on the subject.[60] Lacking the technical know-how and tools to perfectly reproduce this clustering analysis exactly, I conducted a cluster analysis of my own in R using Quanteda. To do so, I first cleaned the text by stripping whitespace and selected for unique express comments only, as this is mostly only used by individuals rather than organizations. I next created a corpus in Quanteda, cast to a document-level matrix, and then compared the Euclidean string distances of all unique pieces of text to all other unique pieces of text. I ran a hierarchical cluster analysis on these data, and cut the tree at the fourth level; I then transformed these data into a dataframe, and filtered to

[60] Kao, *J2kao/Fcc_nn_research*.

only include a frequency of groups higher than half a standard deviation above the mean as mass-comment-like posts. This filter seemed to give the best clustered results to catch the set of unique pieces of text that were just variations of form comments without detecting any false positives. It should be noted, though, that there are likely a few similar pieces of text that were not detected by this process due to their numerical infrequency. Since these did not account for comments which were already flagged as repetitive when filtering by unique, I additionally marked any text that appeared more than twice in the data as a form comment. This was to prevent comments with short strings of text, like "this sucks" or equivalent, from being flagged as true mass comments; one could imagine two completely distinct individuals with no link posting identical, short comments like this. Finally, after joining this back to the text to classify each comment, I found that approximately 98% of comments were mass comments with very little if any unique text added.

This left 2391 unique comments in the dataset, which I attempted to code using the 2015 Lexicoder Sentiment Dictionary embedded in Quanteda. However, when I attempted this, it was abundantly clear very quickly that the built-in sentiment dictionary was misclassifying almost all pieces of text in the wrong direction. As such, I went with the technique of hand classifying the comments, which was already used by two teams of researchers studying the Net Neutrality dataset (Kao, 2017).[61] Since the size of the unique text corpus was small, I decided to hand code every comment. Most comments were clearly for or against a policy, but as a rule, generic anti-government or anti-Google comments that did not specify opposition were coded as such. This was due to the frequency by which those opposed to the policy used anti-government and anti-Obama comments to justify their opposition, as well as stating that the policy was crony capitalism because it allowed companies like Google to continue privacy violations while banning broadband providers. In general, if comments talked about privacy in a positive light — for example that the FCC should ensure their privacy online — they were classified as favoring the policy. The vast majority of coded comments were

[61] Jeff Kao, "Fighting Disinformation Using NLP," Slide 23.

straightforward in their coding, and the full set of coded comments is available in Appendix 1.

Ecological Fallacy Concerns

Since the use of group level demographic and election data will be used here to deduce demographic correlations at the individual level, it is useful here to discuss ecological fallacy concerns — something that this paper cannot fully avoid. I attempt to allay these concerns somewhat by attempting to reproduce my findings with two readily available voter files from North Carolina and Florida (mentioned below), but due to voter file data limitations, they do not contain information on income and thus cannot be used to verify or replicate those findings. We therefore must especially consider the impact of differentially distinct groups within each census block tract by income that may skew correlations in some way. While there are many possibilities, an example of a potential concern might be if, say, unemployed individuals in higher income neighborhoods are more likely to be the adult children of employed parents relative to lower income neighborhoods. This may mean that these unemployed individuals may have more time to leave an original comment in the first place relative to unemployed individuals in lower income areas, thus positively biasing the correlation. Fortunately, given the tight-knit nature of census block groups — each of which contains an average of only one thousand people — it is less likely for many neighbors to have *radically* different demographics from each other in general. While this isn't to say the ecological fallacy concern is no longer an issue, this should somewhat help minimize differential distribution within groups. Unfortunately, barring the use of a national voter file or similarly expensive commercial product, there is no real way to avoid these concerns.

Matching on Voter File

In order to combat ecological fallacy concerns at the census block tract level, as well as take advantage of the individual level dataset, I chose to match individual commenters to the voter files of both Florida and North Carolina. These states were chosen out of necessity for free and available data — North Carolina has a publicly voter file, and the Florida voter file was graciously donated by Professor Fowler — but should also provide diverse representations of potential commenters. Here, both Florida and North Carolina are pretty close to being swing states, providing a semi-equal set of individuals affiliated with different political parties, as well as a diverse set of age and ethnic groups (relative to some other states).

I utilized a two-step matching process for joining Florida and North Carolina voter files to individual level FCC data. First, individuals were matched based on an identical name and street address, which worked on around 15% of commenters from both states. To extract more matches from the dataset, I additionally matched based on zip code, name, and listed city. Since this could lead to numerous false positives for common names, street address data from both the voter file and FCC comment dataset were compared using cosine string distance; matches with a distance of less than 0.2 were kept, and all others were dropped. Additionally, for ease of analysis, multiple registrations under the same name and address were dropped, although this was quite rare. This resulted in a match rate from both states of about 30%. In order to compare these matched voters to the general population of the state, I pulled a random sample of 60,000 voters from the voter file in each state; while direct comparison was possible, computer memory limitations prohibited examining the entire voter file. It is important to quickly note that this comparison can only be made between the population on the voter file, as different demographic groups have different propensities of registering to vote in the first place.

Data Analysis

Data Overview

In all, there are 155,865 successfully geocoded public comments submitted by individuals in this dataset, down from an original 275,000. These comments have been filtered for duplicate submissions — that is, the same person submitting the same text more than once. Since the aim of this paper is to discuss who submitted public comments, as well as how they feel about the Order, this filtering necessarily must exclude duplicate messages from what Shulman (2009) calls "Plebers," who one can regard as individuals that submit the same exact public comment more than once. While 11,000 addresses submitted more than one comment to the FCC — with a median message frequency of 2 and an average frequency of 2.3 messages— of this repeated-submission group, only 214 addresses submitted identical messages. Though this may be one measure of strength of opinion, this paper does not focus on it as such. As can be seen in Figure 1, the 155,000 unique public comments in the dataset from individuals include literature-relevant demographic indicators.

Fig 1: FCC Broadband Privacy Order Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Sentiment	155637	0.052	0.222	0.000	0.000	0.000	1.000
Mass Comment	155865	0.983	0.128	0	1	1	1
Proportion of White Individuals	155865	0.708	0.234	0.000	0.578	0.885	4.127
Proportion of Male Individuals	155865	0.461	0.059	0.000	0.438	0.490	2.985
Median Household Income ($1000s)	153569	60.19	29.625	2.499	39.712	74.014	250.001
Median Age	155865	38.6	7.697	0.000	33.400	43.000	86.000
Has Broadband	155863	0.987	0.112	0.000	1.000	1.000	1.000
Trump Two-Party Vote Share	155565	0.517	0.224	0.000	0.361	0.691	1.000

Here, the variable sentiment represents the support or opposition of an individual to the Order's proposal; as a clarifying aside, it does not reference any sort of expressed tone or emotion. It can take either a value of a 0, which represents advocating against adopting the Order, or 1, representing a comment which advocates for adopting the Order. As is evident in Figure 1, the average sentiment of commenters within the dataset is approximately 5.2%

in favor of adopting the policy. Next, the variable Mass Comment represents whether a comment from an individual has been flagged as a mass comment, as per the methodology. It can take a value of 0, meaning the comment is original, or 1, meaning the comment is a mass comment. On average, 98.3% of comments in the dataset are mass comments. Of note is that, although there are 155,565 comments in the dataset, the dataset is still slightly dirty; N values for each have been included, as almost, but not every single comment, has all variables filled out.

Other variables include the proportion of white voters, which represents the Census-calculated proportion of the population of the block group of the commenter that is white, the proportion of male commenters, which represents the Census-calculated proportion of the population of the block group of the commenter that is male, as well as median household income, which represents the median household income in thousands of dollars of the block group in which the commenter resides. Similarly, median age represents the median age of individuals in the block group in which the commenter resides. Additionally, has broadband is a binary variable derived from the FCC's broadband coverage database — a 0 represents that the census block group where the commenter lives does not have sufficient internet speed (25/3 mbps) for broadband, and a 1 does. Finally, trump two-party vote share represents the average two-party vote share for Donald Trump in the block group of the commenter as a proportion. Since precincts may overlap with census block groups, it was simpler to tabulate the average of precinct level results where the commenter resides. However, this does mean that the true partisan leaning by area may not be accurately represented in the dataset; on average, though, these values should prove approximately correct.

Several findings from these summary statistics alone ought to jump out. First, compared to the general population — which had a two-party 2016 vote share of approximately 48.9% for Donald Trump — the individual commenter dataset containing census block group level data skews nearly three percentage points more Republican with a two-party vote share of 51.7%. Additionally, at 71%, the average proportion of white individuals in a census block

group is less white than the population as a whole, at 76.8%.[62] The dataset also skews more female, with an average census block proportion of 46.1% men, compared to 49.2% men in the general population. Similarly, in June of 2016, the FCC reported that only 92.7% of the population had access to a broadband connection; these data skew more connected, at 98.7% of all commenting census block groups having access to broadband using the same dataset.[63] The dataset also skews slightly older on average, with a median age of 38.6 in commenting block groups, compared to the median national age of 37.2 in 2010.[64] Finally, the dataset skews richer than the average American, with the average median household income in 2010 (the Census dataset vintage) at $50,046 in the general population, compared to $60,194 in the dataset.[65] Both incomes are in 2010 dollars.

Fig 2: Public Polling Results of Internet Privacy Order

Category	Group	Support (%)	Oppose (%)	Not Sure (%)	Total (%)	Unweighted N
Sex	Male	73	12	15	100	426
	Female	70	12	18	100	570
Age	Under 30	65	10	24	100	171
	30-44	71	14	15	100	291
	45-64	72	11	17	100	339
	65+	77	11	11	100	195
Race	White	75	11	13	100	732
	Black	63	10	27	100	98
	Hispanic	48	23	29	100	98
Party ID	Democratic	72	15	12	100	362
	Independent	71	8	22	100	286
	Republican	72	15	14	100	248
Household Income	Under 50k	66	12	21	100	460
	50-100k	78	9	13	100	272
	100k+	74	13	7	100	126

Source: Huffington Post/YouGov Online Privacy Poll

Figure 2 presents the results of a public poll of 1000 US adults on the Broadband Privacy Order in March of 2017, prior to its Congressional repeal. When informed about the text of the Order — namely, that it banned broadband companies from sharing customer

[62] "U.S. Census Bureau QuickFacts."
[63] "FCC Fixed Broadband Deployment."
[64] US Census Bureau, "Median Age Doesn't Tell the Whole Story."
[65] American Community Survey, "Median Income in the Past 12 Months."

information without permission, and that Republicans wanted to overturn the rule before it went into effect — 71% of respondents indicated they wanted the ban to go into effect, whereas only 12% wanted it to be overturned by Republicans. As can be seen in Figure 2, these results are mostly robust to sex, age, party, and household income; they are also robust to white respondents, but not to Black or Hispanic respondents. It should therefore be surprising that only an average of 5.5% of comments from block groups were supportive of this policy, given its wide demographic appeal. Even when we consider the unrepresentative demographic nature of the dataset relative to the population, the minimal support for the Order is extremely noteworthy. Broadly, these polled demographic groups — sex, age, race, party ID, and household income — will motivate variables used to analyze and compare sentiment to individual comments in the dataset.

Mass Comments — Not Bots, But Astroturfing

Since 98.3% of the individual comment dataset is a mass comment — nearly identical to the Cuellar (2004) finding — it is necessary to establish that these mass comments are the result of manufactured salience, rather than bots. To do so, it is useful to examine the three most frequently occurring mass comments, as found in Figure 3.

After an extensive investigation into the source of these mass comments, I believe that, while it is impossible to determine for certain, I have found two distinct sets of evidence that at least the first set of mass comments are human-made and human-submitted form letters, rather than bot-generated. The first is that, in five cases, comments nearly identical to the first mass comment template were submitted that began with "Suggested Comment Language." In two of these cases, these ended with the phrase "Sincerely, YOUR BRIEF COMMENTS."[66] Naturally, this sort of rare template submission mix-up could only be human-caused; clearly individual commenters simply forgot to remove or change the phrases.

[66] Ihkas Riapos.

Figure 3: Top Three Most Frequent Mass Comments

Message (Cleaned, Annotations My Own)	**Frequency (Estimate)**
Dear Chairman Wheeler, You are doing absolutely nothing to protect our online privacy. In fact you are using the power of government to create two sets of rules while protecting certain hand-picked companies from any privacy rules whatsoever! Your privacy plan does not regulate certain business practices but rather selectively regulates those business practices by certain companies but not others. This is the absolute worst example of industrial policy because it has government rather than the free market pick winners and losers in the marketplace.	∼60,000
Dear Chairman Wheeler, The FCC's proposed privacy plan does absolutely nothing to protect Internet users' privacy. Since by your own statements the FCC will do nothing to change the privacy guidelines of the so-called edge providers including the two search and social network giants that by themselves account for 64% of all digital advertising revenues your plan simply won't increase my online privacy. Instead you are using the power of government to pick winners and losers. Enacting rules to block one set of companies from engaging in the same business activities as another favored set of companies creates an uneven playing field stifles competition and is an example of industrial policy at its worst. Your plan doesn't expand Internet users' online privacy rights. It merely restricts any potential competitors from threatening this White House's corporate allies.	∼51,000
Why are you going out of your way to protect Google? Could it be because Google is Obama's biggest corporate patron? Could it be because Google and the federal government work so closely on so many projects?	∼25,000 Note: Includes similar, but non-identical, comments.

Thus, it should be quite clear from these "smoking gun" comments alone that the first set of comments is a real mass comment, rather than simply bot activity and spam. However, where the first set of comments comes from is also a fascinating question, and provides the second set of proof for humans, rather than bots, sending messages. While it is difficult to perfectly confirm the source of the first set of mass comments for certain, the now-defunct, very-likely-astroturfing advocacy website called Protect Internet Freedom is a probable origin for all three of the most common messages. While the website has been taken down in the intervening five years, the Internet Archive has helpfully retained a somewhat-broken copy from 2017.[67] From this, it is possible to deduce several pieces of information.

First, as its own website claims, on April 27th, Protect Internet Freedom submitted its first comment on behalf of a member onto the ECFS docket. While the attached confirmatory screenshot is not available due to backup problems, there was only a single comment received by the FCC on April 27.[68] Not so coincidentally, its text is extremely similar in message to the third most common set of comments directly complaining about Google, for which there are around 25,000. Interestingly, these are found in sets of a few thousand identical

[67] "Blog - Protect Internet Freedom."
[68] Lois Grebosky.

mass comments, with the exact text complaining about Google being let off easy changing from set to set. On its own account, by May 13, over 2000 member comments were left on the docket, which is the *vast* majority of comments submitted by this time.[69] Clearly, then, Protect Internet Freedom was gathering member comments — almost all of which after the first few dozen were identical — and posting them on behalf of members to the FCC docket.

However, the link to manufactured salience is made clearer with the content of a page calling interested members to action. It contains phrases like "sometimes it's hard to tell where Google ends and the Obama administration begins," as well as "while regular Internet Service Providers will operate under the most stringent privacy rules on the Internet, Google will remain free to collect and sell your personal information to anyone they want." Finally, the phrase in the most common message complaining about Google and Obama is directly stated on the website: "Why are you going out of your way to protect Google, Obama's biggest corporate patron?"[70] While these phrases and others clearly link Protect Internet Freedom as the source of the third-most-common set of messages, it is also very closely linked to the other two sets by its comment submission page.

The comment submission page contains two objects: a form to submit comments on one side of the page, which asks for filer name, email, address, and brief comments; and several paragraphs explaining Protect Internet Freedom's point of view on the other. From investigation, the site that the form to submit comments links participants to — called FCC Connect, at connectfcc.com — attempts to fool participants into believing that it is a real portal to the FCC.[71] While the site is also long-dead, it has been used by another, similarly-named website like Free Our Internet to submit similar comments opposed to other policies with the same embedded form. These also criticize Obama, using much the same language, though on different policy issues. More importantly, though, a capture from Free Our Internet from the same time period as the Protect Internet Freedom capture (May 2017)

[69] "Throttled!"
[70] "Tell the FCC's Chairman Wheeler: Your Privacy Rules Are a Farce!"
[71] "FCC Connect."

also contains pre-filled form text; the form format, inputs, and even color is identical besides it.[72] Thus, I am led to believe that Protect Internet Freedom once contained pre-written form text in its comment submission form; the Internet Archive likely did not correctly capture the Protect Internet Freedom page, which had already happened with other pages on the site. Unfortunately, this means that definitive proof of the source of the two most frequently submitted mass comments cannot be found.

However, the wording of each of the two most frequent mass comments almost exactly match phrases found in the issue messaging part of the page. For example, the phrase "[h]e is using government to create two sets of rules, while protecting his friends from any privacy considerations whatsoever!" almost exactly matches one of the first sentences of the most popular mass comment, which reads ". . . using the power of government to create two sets of rules while protecting certain hand-picked companies from any privacy rules whatsoever!"[73] Similarly, the next sentence of the issue messaging section reads "[t]he FCC is not planning to regulate business practice uniformly, but rather to selectively regulate that practice only for certain companies and not others," which is barely any different from the next line of the mass comment, which is "... does not regulate certain business practices but rather selectively regulates those business practices by certain companies but not others."[74] Other, shorter phrases that are very similar to the most frequent mass comment are also used, but I leave this as an exercise to the reader. As a final link to Protect Internet Freedom, two comments begin with the line:

> "Note: You are filing a document into an official FCC proceeding. All information submitted including names and address will be publicly available via the web. There was a problem with your submission. Errors have been highlighted below. Information NAME OF FILER* EMAIL ADDRESS Address ADDRESS* ADDRESS LINE 1 ADDRESS LINE 2 CITY STATE [. . .] ZIP Comments

[72] "Take Action - Free Our Internet."
[73] "Tell the FCC: Let Me Have Some Privacy."
[74] Ibid.

SUGGESTED COMMENT LANGUAGE . . .”[75]

While this may appear to be gibberish, it is in fact exactly the text of the submission form on the Protect Internet Freedom page. Unsurprisingly, this is immediately followed by the full text of the most frequent mass comment.[76]

Interestingly, the second most frequent set of mass comments are less obviously linked to phrases already on the page, but are nonetheless clearly similar in message, tone, and occasionally phrase with both the page and the most frequent set of mass comments. It is distinctly possible that there is another, but now-broken page on Protect Internet Freedom that has similarly near-identical messaging. Regardless, this is clear proof that Protect Internet Freedom was the source of most, if not all, of the form letters, and that it additionally engaged in manufactured salience messaging tactics by recommending “messages and words” to individuals submitting comments, as Jordan and Watson (2019) lay out.

For true proof of the Jordan and Watson conception of manufactured salience, though, there needs to be somewhat-clear proof that Protect Internet Freedom is an astroturfing organization rather than a grassroots organizing campaign. Unsurprisingly, there is very little information to be found about the organization online, and the Our Team link on the website for Protect Internet Freedom contains very little substantial information, other than implying that all of its seven staff and advisory board members worked for conservative candidates and causes. Fortunately, a search for Protect Internet Freedom on the IRS’ Tax Exempt Organization search provides slightly more information, mentioning Protect Internet Freedom’s principal officer is John Henke, and that it has a mailing address to a nondescript P.O. Box in Arlington, Virginia. Unfortunately, since the organization reported less than $50,000 in receipts in 2016 and 2017, it is impossible to see where its funding came from.[77]

However, John Henke’s background provides some clues as to the true source of funding

[75] Alan Douglas.
[76] “Tell the FCC: Let Me Have Some Privacy.”
[77] “Protect Internet Freedom Foundation.”

— and motivation — for Protect Internet Freedom. Henke is a former political blogger turned internet strategy and communications consultant, and has been the Executive Director of a digital strategy organization accused of helping corporations astroturf online policy conversations.[78] In the past, he has also worked for a group called Arts+Labs in 2009, which purported to be an organization that advocated for digital artists and other new-media-related businesses, but instead (allegedly) attempted to persuade individuals and the FCC that the entertainment industry's preferred internet-related policies were sorely needed and best for the country. Naturally, the site partnered with the large entertainment industry companies and lobbying organizations, including Viacom, NBC Universal, AT&T, BMI, Verizon, Microsoft, SGA, and ASCAP.[79] While it is again impossible to say for certain what organizations Protect Internet Freedom partnered with or was funded by, its Principal Officer — suspiciously not listed on the Our Team section of the site — provides compelling clues that it is likely not a grassroots organization after all. While only circumstantial, this more than hints that Protect Internet Freedom and its associated mass comments meet the criteria to be truly considered manufactured salience.

Block Group Sentiment

Although it should be clear that the vast majority of all comments in the dataset are the result of manufactured salience, as discussed in the literature review section, these comments are still sent by individuals who do so to express their own identities (Jordan & Watson 2019). Since these are still valid comments, it is necessary to look at demographic factors that correlate with overall block group level sentiment before investigating the demographics that correlate with sending mass comments.

To do so, I ran an OLS regression model, the results of which are shown in Figure 4. Here, the dependent variable is block group sentiment which, as noted previously, can take

[78] Phillip Dampier, "Broadband Usage Caps."
[79] Ibid.

Figure 4: Block Group Level Order Sentiment	
	Sentiment
	(1)
Trump Two-Party Vote Share	-0.279***
	(0.003)
Proportion of White Individuals	0.113***
	(0.003)
Proportion of Male Individuals	-0.024**
	(0.010)
Median Household Income ($1000s)	0.001***
	(0.00002)
Has Broadband	-0.023***
	(0.005)
Constant	0.091***
	(0.007)
Observations	101,961
Adjusted R^2	0.074
Residual Std. Error	0.211 (df = 101955)
Note:	**p<0.05; ***p<0.01

on either a 0 or a 1 for each individual comment. While all chosen demographic factors were statistically significant at the 95% level, the most substantively significant variables in the first model were the average proportion of two-party Trump vote share in the block group the individual commenter resides in, followed by median household income in thousands of dollars of the block group the individual commenter lives in, the proportion of white individuals in the block group the individual commenter resides in, and finally having broadband in the block group where the individual commenter lives. First, holding all else constant, when moving from the 25th percentile of block groups for Trump two-party vote-share to the 75th percentile, the average comment sentiment in a block group decreased by 9.2 percentage points. That is, when moving from the 25th percentile to the 75th percentile (36% to 69% of two-party vote share), we expect that an additional nine percentage points of comments in a block group would oppose the Order.

Next, when moving from the 25th percentile to the 75th percentile of the block group of a commenter for median household income in thousands of dollars — a movement from

$40,000 to $74,000 — we would expect that the average sentiment of an individual comment would increase by 3.4 percentage points. Similarly, moving from the 25th percentile to the 75th percentile of commenters' block groups for the proportion of white individuals in a block group (from 57.8% to 88%) was correlated with a 3.5 percentage point increase in the average sentiment of an individual comment. Finally, moving from a block group without broadband to a block group with broadband was associated with a 2.3 percentage point decrease in the average sentiment of an individual. While the proportion of male individuals in a block group was statistically significant, the small interquartile range meant that its expected change was essentially negligible. Similarly, age, while not included in the above models, was also statistically but not substantively significant when included. A second model without either of the least-substantively-significant variables included, not shown, was also analyzed to check robustness.

A comparison to the initial public poll, shown in Figure 2, provides some interesting results. Most interestingly, the Trump two-party vote share finding — namely, that an individual in a more-Republican block group is correlated with an increase in opposition to the Order — does not match the Party ID results of the poll, which finds no difference in support between Republican and Democratic voters. While it was never expected to, this finding implies that commenters in this set of public comments are unrepresentative of the overall adult US population surveyed in the poll. On the other hand, the results for both the proportion of white voters and the median household income of the block group an individual resides in do broadly match up with the results of the poll. In the former's case, white voters appear to know more about the policy than non-white voters, some of whom are also more likely to oppose the Order when they do. While the set of individuals who did not know about the Order and thus could not comment is not known, this decrease in support from individuals in less-white block groups is reflected in the data. In the latter's case, the poll demonstrates a clear increase in support among higher household income adults when compared to lower household income adults; this is also reflected in the OLS model above.

The lack of substantive significance in comments submitted from more-male block groups also confirms the results of the poll, which found little to no meaningful difference between sexes. Finally, the finding that moving from a comment in a block group without broadband to a comment in a block group with broadband is, holding all else constant, correlated with a decrease in sentiment seems to disagree with the Singel (2018) finding. Whether this is simply due to a change in issues or a broader phenomenon within all public comments is a topic for future research.

Frequency of Comments

Similarly, investigating correlations between comment frequency and literature-motivated demographics groups can reveal differences in the likelihood of commenting for block groups where there is at least some salience on the Broadband Privacy Order issue. Although ideally this analysis would have been conducted at the individual level, the lack of a large number of individuals submitting more than one comment requires aggregation to the block group level.

Since this dataset does not record block groups where no respondent commented — and only represents around 50% of all possible block groups — these correlations can only reveal differences in frequency between block groups that have already commented. While it is possible that these results may hold upon inclusion of the other 50% of block groups, the differences between the general population and block groups in the dataset suggests that taking this assumption would be misplaced. As such, the task of comparing frequency against the entire population is left for future studies on the topic.

In order to tease out commenting rate differences amongst block groups that had commented on the FCC dataset, the data were aggregated into block groups, where each row represents one block group. I then ran two OLS regressions, the results of which are described above in Figure 5. The dependent variable, count of responses in a block group,

Figure 5: Block Group Level Comment Frequency

	Count of Responses	
	(1)	(2)
Trump Two-Party Vote Share	0.539***	0.532***
	(0.020)	(0.019)
Proportion White Individuals	-0.321***	-0.315***
	(0.020)	(0.019)
Proportion Male Individuals	-0.884***	-0.877***
	(0.062)	(0.059)
Median Household Income ($1000s)	-0.00003	
	(0.0001)	
Has Broadband	0.216***	0.217***
	(0.028)	(0.028)
Median Age	-0.003***	
	(0.0005)	
Constant	1.655***	1.647***
	(0.041)	(0.040)
Observations	102,030	103,697
Adjusted R^2	0.012	0.012
Residual Std. Error	1.052 (df = 102024)	1.050 (df = 103692)
Note:	**p<0.05; ***p<0.01	

necessarily has to be at least equal to one to be included in this dataset. On average, each block group in the dataset has just over one comment submitted on the Order, and ranges from one comment to a maximum of 18, with a median value of one. All but the median household income was significant at the 95% confidence level, which is displayed in model 1 for convenience. In rank order, the most substantively significant variables were whether a block group has broadband access, followed by average Trump two-party vote share in the block group and the proportion of white individuals in a block group. All other variables had a substantive interpretation of an increase from the 25th percentile to the 75th percentile of an increase or decrease of less than 0.09 additional comments. Model 2 contains only these substantively significant variables, but also includes the proportion of male individuals in a block group due to its surprisingly large effect on the proportion of white individuals when excluded.

As can be seen in Model 2, the most substantively significant demographic variable was moving from no broadband to broadband in a block group, which is correlated with an average increase of 0.217 comments in a block group. A similarly large effect size can be found when moving from the 25th percentile to the 75th percentile of average Trump two-party vote share in a block group, which is correlated with an average expected increase of 0.18 comments in a block group. Following this, the most substantively significant variable is the proportion of white individuals in a district, which, when moving from the 25th percentile to the 75th percentile of block groups, is correlated with a decrease of approximately 0.10 comments.

While not completely comparable to the general population, these results should still be surprising, given the existing literature on differential participation. From Olsen, Galloway and Ruth (2009), as well as Farina (2011), we would expect older and wealthier individuals to submit relatively more public comments, neither of which shows up in these models of increased participation; similarly, although the sign is correct in both models, we would expect a larger difference in comment frequency between male and female individuals, as mentioned in Bryer (2013). Unsurprisingly, though, the model does provide evidence that the digital divide impacts commenting frequency, as we would expect from Remaley (2020). This digital divide-caused difference is also correlated with an increased proportion of white individuals. Given the sign and substantive significance of the proportion of white individuals in a block group, the model results are thus quite perplexing, as the frequency of comments would be expected to increase, rather than decrease. While this could be simply a case of Simpson's Paradox or another ecological fallacy concern, it certainly indicates that a true individual-level analysis on race is warranted.

Mass Comments Within Block Groups

It is also possible to tease out correlations between individuals submitting mass comments and various demographic characteristics, which will help test claims made by Schlosberg, Zavestoski, and Shulman (2007). This analysis can, apart from previously mentioned ecological fallacy concerns, also reveal which identities — demographic groups — are being targeted by Protect Internet Freedom in the first place. This analysis will only return meaningful results if the vast majority of mass comments have the same classified sentiment and were sent by the same organization. While I discuss the latter in an earlier part of this section, the former is also true, as over 95% of all mass comments in the dataset were classified as against the order. While this means that any results will necessarily also be measuring a small amount of positive mass comments not from Protect Internet Freedom, thus increasing variation, the correlations should roughly only include targeted groups.

Figure 6: Block Group Level Mass Comments

	Mass Comment	
	(1)	(2)
Trump Two-Party Vote Share	0.045***	0.043***
	(0.002)	(0.002)
Proportion White Individuals	-0.024***	-0.022***
	(0.002)	(0.002)
Median Household Income ($1000s)	-0.0002***	-0.0002***
	(0.00001)	(0.00001)
Median Age	-0.0003***	-0.0003***
	(0.00005)	(0.00005)
Proportion Male Individuals	0.017***	
	(0.006)	
Has Broadband	0.004	
	(0.003)	
Constant	0.990***	1.002***
	(0.005)	(0.002)
Observations	102,030	103,697
Adjusted R2	0.012	0.012
Residual Std. Error	1.052 (df = 102024)	1.050 (df = 103692)
Note:	***p<0.01	

To investigate potential demographic correlations, I once again ran two OLS regression models, which are presented in Figure 6. Here, the dependent variable is whether an individual submitted a mass comment (represented by a 1) or an original comment (a 0). Model 1 contains the results of the regression model with all tested variables; of these, neither the proportion of male individuals in a block group, nor moving from no broadband to broadband access in a block group, were substantively nor statistically significant respectively. Of the variables included in Model 2, in descending order, the proportion of two-party Trump vote share, the median household income of the block group, the proportion of white individuals in a block group, and the median age were substantively significant. Since over 98% of all individual comments are mass comments, substantively significant movements here are quite small, ranging from just over one percentage point to 0.3 percentage points of movement.

First, moving from the 25th to 75th percentile of commenters in two-party Trump vote share is correlated with a 1.5 percentage point increase in the probability of an individual submitting a mass comment; moving from the 25th percentile to the 75th percentile in median household income also correlates with an increase of 0.7 percentage points in probability. The proportion of white individuals in a commenter's block group has a similar magnitude when moving from the 25th to 75th percentile of commenters' block groups, with an expected average 0.7 percentage point decrease in the probability of submitting a mass comment. Finally, moving from the 25th to 75th percentile of median age of commenters was correlated with an expected decrease of 0.2 percentage points in the probability of submitting a mass comment.

These findings partially confirm and partially disagree with the conclusions made by Schlosberg, Zavestoski, and Shulman (2007). While Schlosberg et al predicts that those submitting mass comments were more likely to be women, the results of this regression fail to reproduce this, showing no difference between the two sexes at the block group level. This is possibly due to a small effect size, or possibly due to ecological fallacy concerns, but points towards the need for further analysis. Unfortunately, the dataset does not contain

metrics to test education predictions, but the above model does confirm that commenters in block groups with lower incomes are more likely to submit mass comments than commenters in block groups with higher incomes, confirming the Schlosberg finding. However, it is also possible that the above findings are simply due to intentional targeting from Protect Internet Freedom, which certainly would confound the results. For example, one could imagine that, while women were more likely to submit mass comments, Protect Internet Freedom targeted men, leading to a negative bias in the results. From the difference in effect size magnitude and the lack of literature to back up variation by party, though, it is clear that intentionally or not, Protect Internet Freedom likely targeted individuals that had Republican-leaning identities (putting aside ecological fallacy concerns). This is backed up by the types of material used by Protect Internet Freedom to attract and convince commenters to post. While this could be somewhat correlated with education, the effect size of two-party vote share, combined with household income, sex, and race variables that would also capture some education effect, seems a bit too large to be the case. Since this study by design cannot separate these effects fully — nor point to causality — future research into identities used by manufactured salience on a broader scale are warranted.

Individual Voter File Level Analysis

In order to provide some verification of findings at the level of the individual — and importantly provide some robustness checks against the ecological fallacy — I matched individual commenters to the voter file in both Florida and North Carolina; more details on the process can be found in the methodology section. It is important to note that, while the 30% match rate for commenters in each state can partially be ascribed to typos or unclean address data, it also reflects that the voter file is itself a selective set. As is well known, the voter file (and voting itself) does not perfectly reflect the population, which for the purposes of this analysis means that conclusions will be limited to comparison between commenters

registered to vote and the average individual registered to vote.[80] Since we cannot assume that all individual commenters in the Order dataset are registered to vote, nor that they register at the same rate as the general population, the findings of this section cannot strictly verify the results from previous data analysis sections. Thus, while these findings are useful for providing additional evidence for — or challenging — the results of previous sections, they can only point to a need for additional research.

Briefly, the voter file dataset contains a set of 60,000 randomly selected individuals on the voter file from both North Carolina and Florida, and additionally contains the matched set of 1630 individual commenters and 3291 individual commenters from North Carolina and Florida respectively. Within these data, the proportion of mass comments in Florida was 99.25%, compared to 98.2% of mass comments in North Carolina. Finally, 98.9% of individual commenters in Florida opposed the policy, compared to 98.7% of commenters in North Carolina. It is also worth noting that the voter file contains information on age, race, sex, and party registration; it does not contain information on income.

[80] Leighley and Nagler, *Who Votes Now?,* Chapter 1.

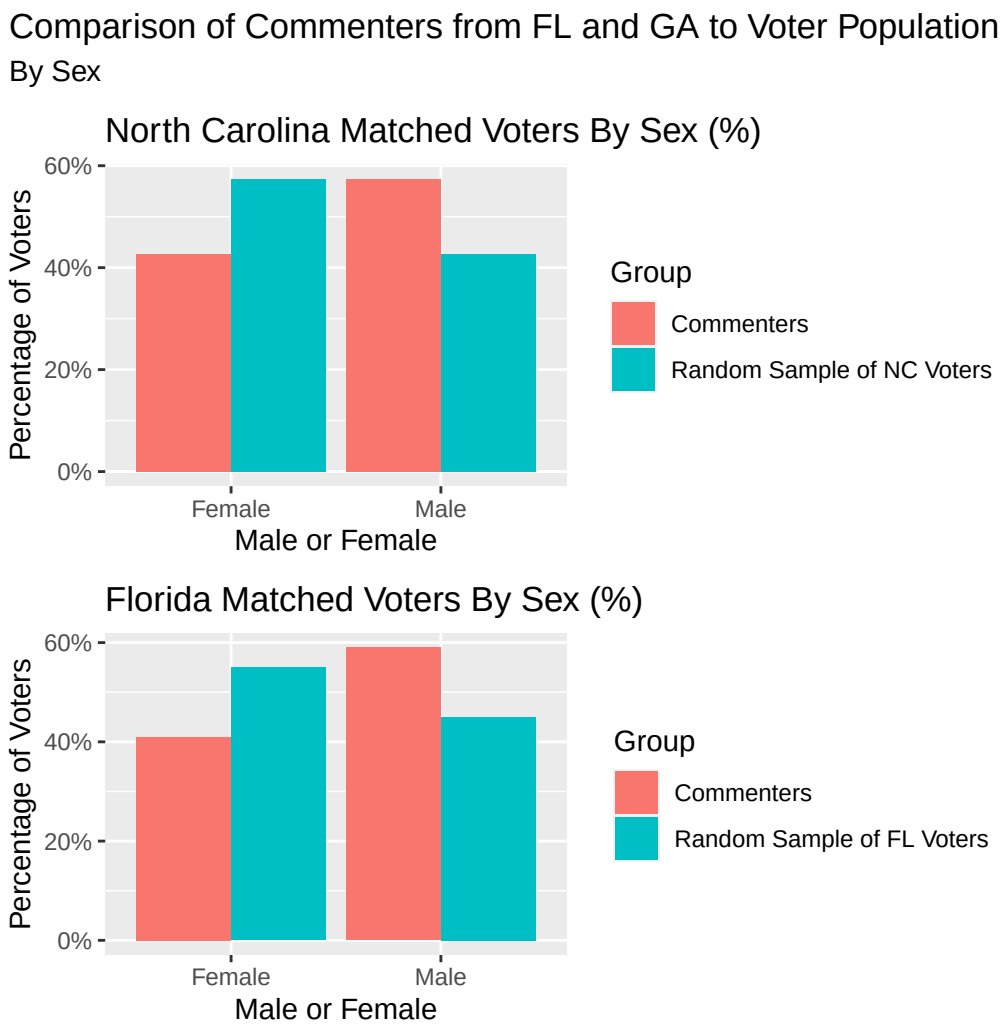

Figure 7: Voter Comparison by Sex

Due to overfitting concerns associated with the very small number of positive and mass comments in the individual commenter voter file dataset, this paper is unable to run regressions within the commenter group on the data; on the other hand, due to the overwhelming number of negative comments relative to positive ones, it is possible to directly compare demographics directly between matched commenters and the voter file. In Figure 7, it is quickly clear that, while women outnumber men by registration in both states, the proportions are almost exactly flipped when it comes to individual commenters; 59 percent of matched commenters are men, compared to just 45 percent of registered voters in Florida.

This is similar to the 57% of commenters versus 43% of registered voters who are men in North Carolina.

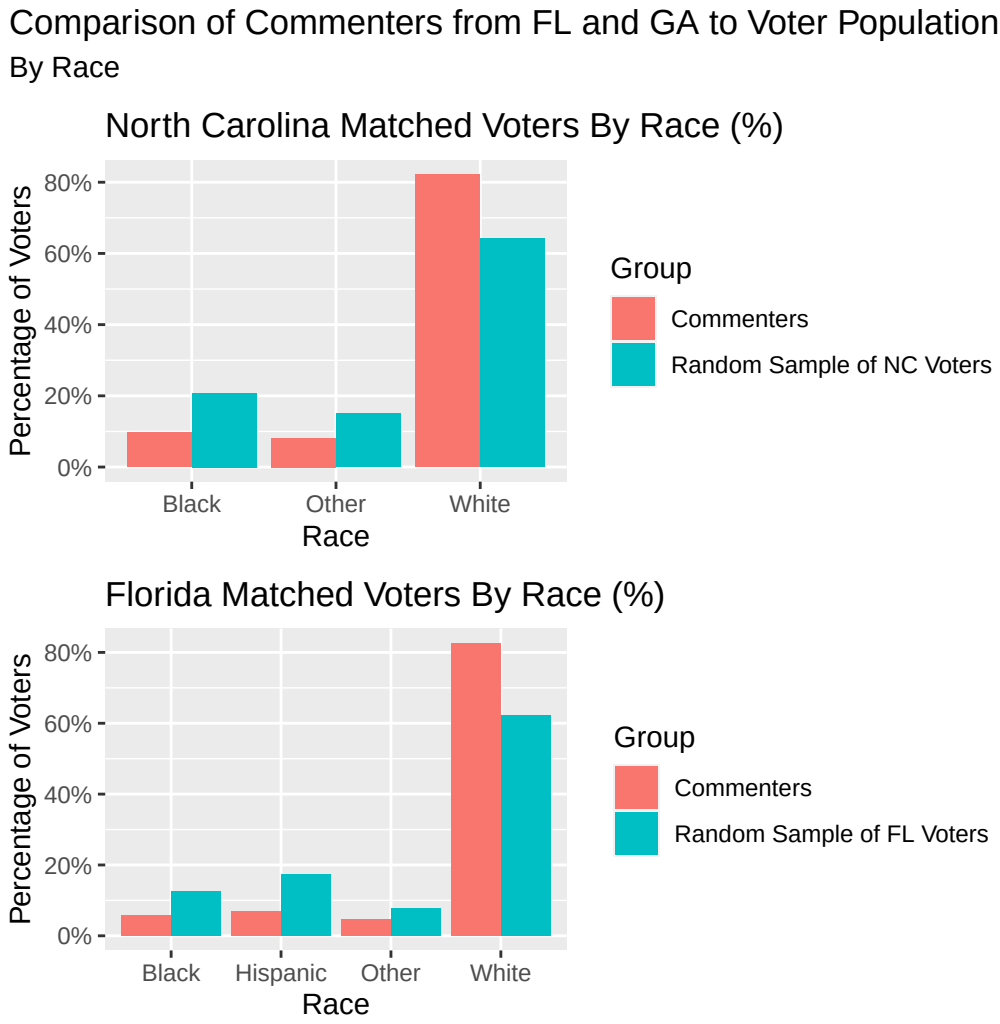

Figure 8: Voter Comparison by Race

There are similarly large discrepancies when investigating race across both the matched commenter and registered voter groups in Figure 8; both Florida and North Carolina have fewer Black, Hispanic, and other non-white individual commenters compared to the registered voter baseline. Both have much higher proportions of white individuals than the set of registered voters — at 82% for both states — compared to the 64% and 62% proportion of white individuals in North Carolina and Florida respectively.

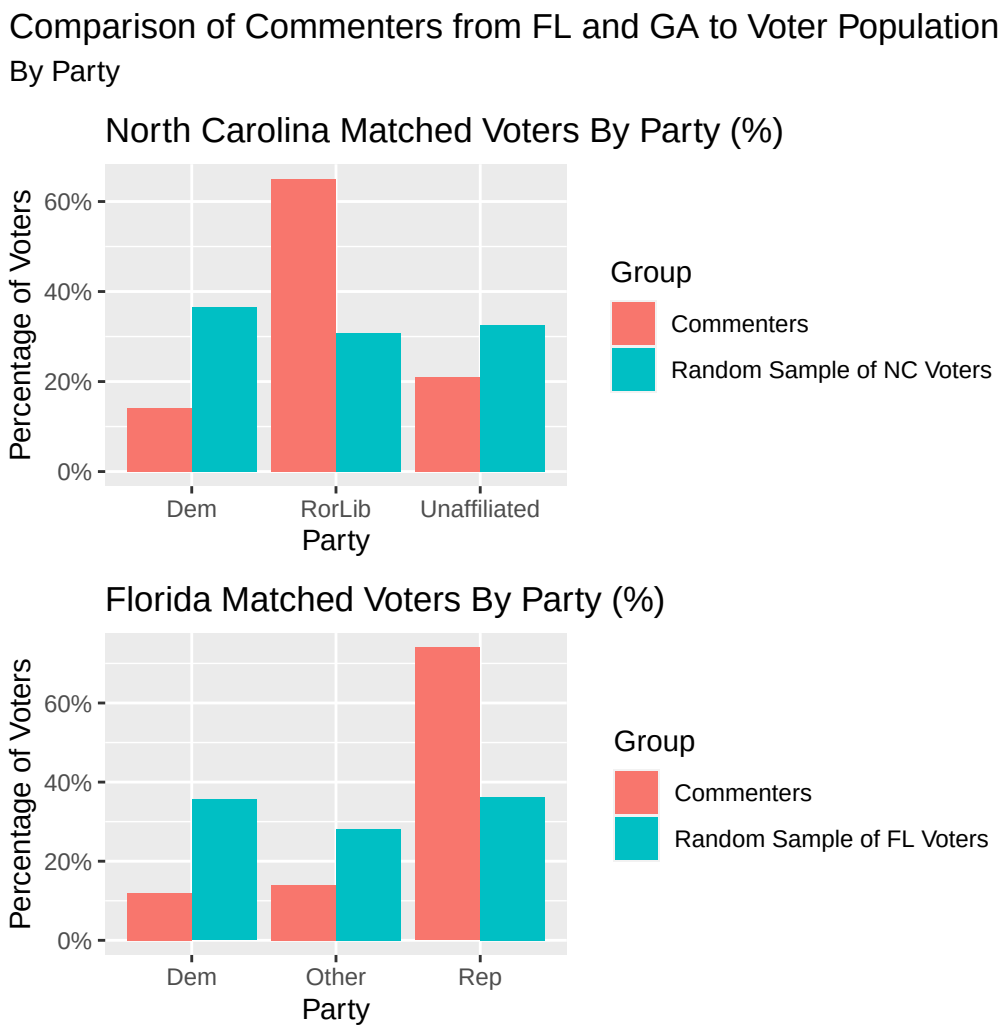

Figure 9: Voter Comparison by Party

Perhaps unsurprisingly given the above differences by race, there are also clear differences compared to the set of registered voters when grouping by party affiliation in both Florida and North Carolina, as seen in Figure 9. In Florida, there are proportionally double the number of registered Republican commenters (74%) than there are in the set of registered voters (36%). Similarly, in North Carolina, there are also more than proportionally double the number of registered Republicans or Libertarians (65%) compared to registered voters (30%). Relatedly, there are proportionally less than a third of individual commenters who are Democrats compared to the registered voter baseline in NC, which is almost exactly

the same in Florida. Since there were only a handful of registered Libertarians, they were lumped into Republicans for ease of comparison to the three columns in the Florida dataset. Additionally, while North Carolina allows for registering as an unaffiliated voter, this option does not appear to be available in Florida — thus, directly comparing between Florida and North Carolina here on non-Democratic or -Republican categories is not possible.

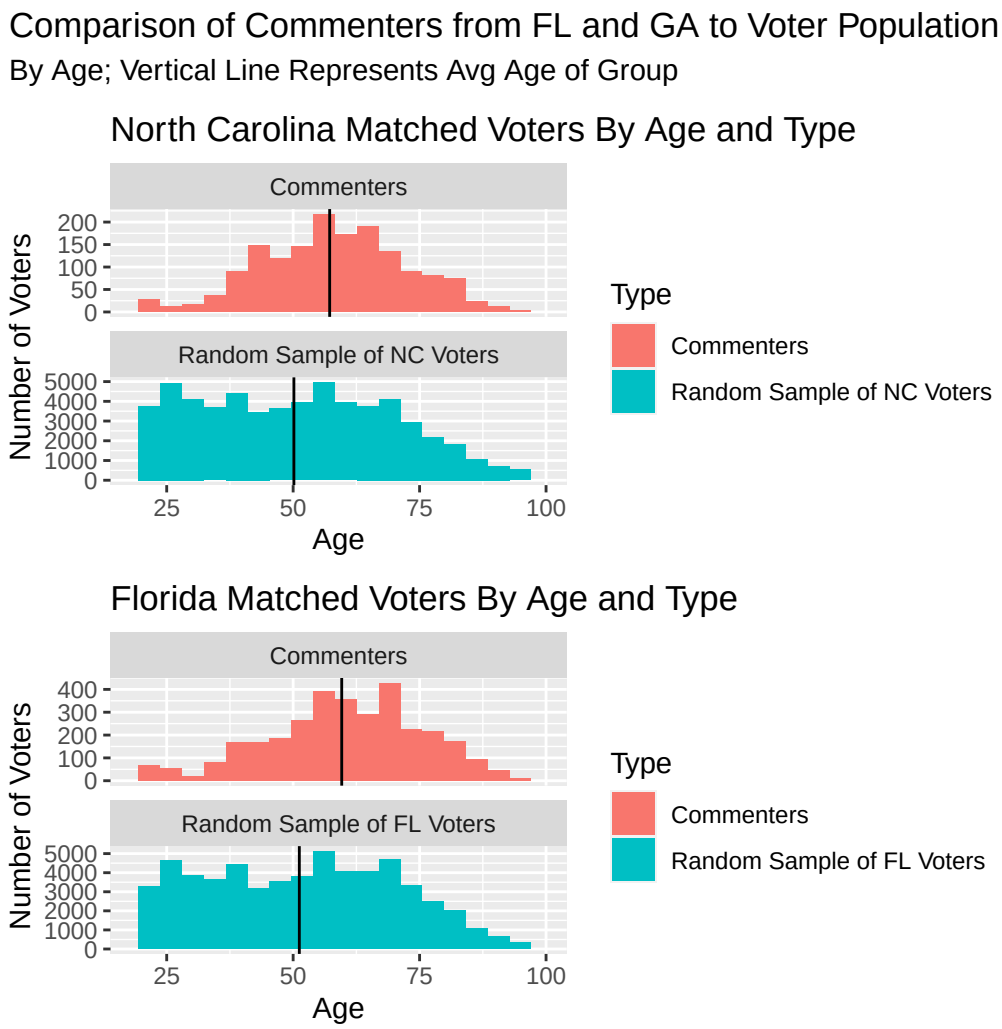

Figure 10: Voter Comparison by Age

Finally, Figure 10 shows clear differences in distribution between individual commenters and a random sample of voters when split by age. Individual commenters are seven and eight years older in age than the average North Carolina or Florida voter respectively, which

appears to be caused by a complete drop-off of individuals aged less than forty years old. The difference in these distributions is clearly very stark, and is very similar in both states.

These individual-level results provide several findings, although they once again are not directly comparable to the general population of commenters or adults for previously mentioned reasons. First, these data seem to confirm the findings derived from block group level ecological data for party vote share and age, as mentioned in the initial overview of the block level data. However, these results run contrary to sex and race findings from the initial summary statistics; from these data, individuals are more male than female, and far more white than non-white. Given that ~99% of all commenters had negative sentiment and submitted a mass comment — indicating that these individuals submitted Protect Internet Freedom messages — these data also qualitatively support the party ID findings from the mass comment section above. Finally, these data reveal a clearly distinct age distribution for the issue compared to the overall population; this seems to go against the findings from the Huffington Post poll that show no meaningful difference in opposition by age, although perhaps the lack of young commenters is simply caused by a lack of knowledge about the policy's existence, which would be supported by the poll.

Reliability of Block Group Level Data

To further validate the reliability of block group level data in making inferences about individuals, I attempted to directly compare the known NC and FL voter file dataset with those derived from block group data. The only available dependent variable shared between the complete voter file dataset (of all voters) and a hypothetical block group dataset of all individuals in either state was a variable denoting whether an individual commented on the Order or not. As such, I could only measure whether group block group data was perfectly predictive of increases in comment probability by demographic group; this would be compared to the known increases in probability of commenting as recorded by the individual

level voter file data. Here, as proportions of the state population, the absolute numbers are irrelevant — only the relative differences in magnitude and direction between the group level probability predictions and true individual level probabilities matter in order to determine whether predictions were an over- or underestimate of the true individual level probability findings. While this was straightforward to calculate using a regression on both sets of voter file data — of which I scaled a smaller set up to reflect the total number of voters in a state for computer memory related reasons — it was less straightforward on block group data that only contained commenting individuals. To maintain the same effective dependent variable, I generated a variable for the proportion of commenters relative to the population of each block group, whether it contained commenters or not, and subsequently weighted for population in the block level variable regressions of both states.[81]

Figure 11: Magnitude Differences Between Prediction and Reality Probabilities By Demographic Variable

Demographic Variable	**Florida**	**North Carolina**
White Individual	2.2x larger	1.3x smaller
Male Individual	3.4x larger	2.4x larger
Age of Individual	1.7x smaller	1.3x larger

Thus, I ran 4 previously-discussed OLS regressions of whether an individual commented on the Order or not against the three variables shared between the two voter files and block group level dataset — race, sex, and age. These variables are not *directly* comparable, as by their nature they were measured within the block group dataset as proportions of white and male individuals, as well as the median age of an individual within the dataset. However, as the goal of this reliability test is to determine whether these group proportions were similar in predictive power to the true individual level data, this is not a problem. As can be seen in Figure 11, when the comparing the magnitudes difference in the probability of commenting given movement from a non-white, female, or one-year-younger individual to a white, male, or year-older individual, the block group level data prediction fared reasonably well in both

[81] Thank you to Professor Fowler for this brilliant idea.

North Carolina and Florida. The specific probabilities of each regression are also available in Appendix 2. In predicting the true increase in probability of commenting for an individual who is one year older than an otherwise-identical individual in Florida, the block group data underpredicts the true probability by a factor of 1.7 times. Similarly, the block group data underpredicted the true increase in probability of commenting when moving from a non-white to a white individual in North Carolina. On the other hand, the block group level data overpredicted the true increase in probability of commenting when moving from a non-white to a white individual in Florida by a factor of 2.2 times, as well as the increase in probability of commenting when an otherwise-identical individual increases in age by one year in North Carolina by a factor of 1.3 times. Finally, in both states, the block group level data overpredicted the increase in probability of commenting when moving from a female to a male by a factor of 3.4 and 2.4 times in Florida and North Carolina respectively. It should additionally be noted that, in all six cases, the sign of the effect was identical.

While the predictions generated by the block group based data do not perfectly match the true probabilities provided by each state's voter file data, they are remarkably similar given significant ecological fallacy concerns. While it appears there may be a tendency for the block group level demographic data to overestimate the true probability by around double, it is also distinctly possible that these regress closer to the mean when averaged over many states and many variables. Indeed, the lack of a consistent over- or underestimate in both the age and race variables demonstrates that this regression to the mean may prove to be true. However, care does need to be taken for results relating to the sex of individual commenters, as these findings do demonstrate a consistent — and large — overestimate of the true probability by sex. In reality, though, this should not be particularly surprising; unlike the clear spatial segregation by age and race evident in the United States, there is very little variation by sex. Thus, the very small random variations in the data may be magnified. Broadly speaking, though, while caution may be needed when considering effect sizes, these results demonstrate the reliability of results derived from block group level data. All results were well within an

order of magnitude of the true probabilities, and the direction of coefficients was consistent across all three studied variables. This provides further evidence for the reliability of other findings based on the same data in this paper, and certainly underscores that the previous findings of this paper likely do truly point to fruitful areas of further research.

Results

The above analysis points to key findings for sex, race, party, income, age, and broadband access. The above findings on sex appear to match sentiment findings from the public poll, disagree with the finding of Schlosberg, Zavestoski, and Shulman (2007) on mass comments, and are more inconclusive on comment submission. First, the results by sex from the sentiment regression clearly match the public poll data when controlling for other demographic variables; there was no meaningful difference between men and women. While the public poll did not control for other demographic variables, these results seem to provide evidence that the lack of a meaningful difference is robust to other variables. Next, the results of the mass comment analysis — also showing no difference in mass commenting by sex at the block group level — provides evidence against the Schlosberg et al finding that women are more likely to submit mass comments. It is not clear from the Schlosberg reading whether we would expect this to be robust to other demographic variables; in either case, these results provide evidence that women are not, holding party, race, income, and broadband constant, more likely to submit mass comments.

Finally, whether these results support a sex difference in submitting public comment at all is more complicated: while the summary stats show a proportional increase of women in the dataset relative to the population, and comment frequency results point to individuals in block groups with a higher proportion of women commenting slightly more frequently, the individual voter file level results show far more men commenting than women. While the frequency results could potentially reflect a difference between the unobserved noncom-

menter women and commenter women — with women who submit comments more likely to submit multiple — they could also be the result of multivariate regression on other demographic factors. The difference in results between the summary statistics and the voter file is somewhat concerning, though; it could point to an unobserved geographic division, simply just be Simpson's Paradox within the dataset, be the result of identity targeting, or most worryingly provide evidence for the ecological fallacy concern mentioned in the methodology section. Thus, these results fail to resolve differences in the Olsen and Bryer literature, and likely point to the need for additional research in this area. It is also, of course, important to note that sex was the least reliable variable when evaluated, and thus the estimated effect sizes within the block group data may be several times too large.

Next, the above findings on race seem to confirm the public-poll-based findings on sentiment by race, as well as potentially disagree with the digital divide-based findings stemming from Remaley (2020). First, the summary statistics and frequency analysis point to both an increase in the proportion of non-white individuals in block groups relative to the general population, as well as an increased number of comments submitted by block groups with more nonwhite participants when controlling for other demographic factors. While the former disagrees with concerns about differential participation by race due to the digital divide — putting aside Simpson's Paradox concerns — in the latter's case, this may be due to controlling for factors like income that also correlate with the digital divide. These may also point to differences in sentiment broadly; the above results show that individuals in more non-white block groups were more likely to oppose the Order, as well as were more likely to submit mass comments. However, these results should be taken with a grain of salt. Matched individual commenters were proportionally much, much whiter than the set of registered voters; this is somewhat tempered by existing differences in voter registration by race, as well as the potential for racial differences by geography. These matched commenter results may also point towards serious ecological fallacy problems, but on net, the combination of the public poll and robustness to other demographic factors by sentiment & frequency seem to

point to higher participation and negative sentiment in non-white individuals.

The findings of this paper are more conclusive when it comes to the party of block groups and individuals. In contrast to the public poll, which found no meaningful difference across Party ID groups, the summary statistics, sentiment, frequency, mass comments, and voter file level analysis point to a clear difference between commenters in less Republican and more Republican block groups that is robust even when controlling for other demographic variables. Additionally, proportionally more individuals commenting are Republicans than would be expected when compared to the population. While the cause of this discrepancy between sentiment measured in the public poll and sentiment of public comments cannot be determined from this paper, it is implied from the literature that Protect Internet Freedom may have tapped into this latent belief with Republican-identifying individuals in order to motivate more comments. Of course, there could also be reverse causation — Republican-identifying individuals were more likely to be susceptible to changing their beliefs toward opposing the Order due to some Protect Internet Freedom-caused factor. Thus, this paper echoes the conclusions of Jordan and Watson (2019) — manufactured salience is clearly a big driver of public comment, and more research into messaging and targeting techniques based on identity are sorely needed.

The above findings on income seem to match public poll data, agree with the Schlosberg findings, and partially agree with the Olsen findings. First, even when controlling for other demographic factors, there is a clear correlation between higher income and positive sentiment, matching public polling data even on a potentially-unrepresentative set of commenters. This is potentially due to the poll-predicted lack of knowledge of the Order among lower income individuals, making participating lower income individuals proportionately lower in sentiment than higher income ones. Similarly, the findings of this paper agree with the initial Schlosberg et al findings that higher income individuals are more likely to submit more original (i.e. less mass) comments. However, the findings of this paper are more complex than what was predicted by Olsen et al; while the initial comment dataset on aver-

age had a higher household income than the average household in 2010, when controlling for other demographic factors the frequency of comments did not appear to be correlated with household income. This lack of correlation may be due to an unobserved difference between lower income non-commenters and lower income commenters.

The above findings are more inconclusive on age. First, the findings imply that the difference in sentiment by age demonstrated in the public poll is not robust to other demographic variables within the set of commenters; there was no significant difference between age and sentiment. However, comparison to the Olsen (2009) and Farina (2011) predictions that older individuals are more likely to submit public comments tells a more complicated story. While the dataset is slightly older, older block groups do seem to submit more comments than younger block groups on average, implying this finding is not robust to other demographic variables among commenters. With this said, the results from an individual voter file analysis are shockingly large; the distribution of commenters by age compared to the population includes barely any individuals younger than age 40, a dramatic difference from the population. This may not be robust to other demographic variables — namely, income and party — but the discrepancy does point to more older registered voters participating than younger registered voters. Whether this holds at the individual level of analysis in other geographic areas, in the set of all adults, and when controlling for other demographic variables is a key question for future research to answer.

Finally, the above findings agree with previous literature on the effect of the digital divide. As predicted, the number of individuals in block groups with broadband access was proportionately much higher than would be expected from the general population; additionally, broadband access was the most substantively significant correlate for commenting frequency at the block group level. Thus, it is quite clear from these findings that individuals in a block group with broadband were more likely to comment than those in a block group without broadband, even when controlling for other potentially confounding demographic factors. Interestingly, while Singel (2019) observed that there was no significant difference

between groups with and without broadband access, these findings reveal a small but significant correlation between broadband access and opposing the Order. As there are no differentially high broadband access rates in Republican states, this does not appear to be due to any similar lurking variable, and thus may prove interesting for future public polls to include this when asking about future broadband-related policies; however, it could also simply be a spurious correlation, or contain another unobserved confounder.

Discussion

Who Participates and How

Although the move towards e-rulemaking was to enhance citizen participation in rulemaking through lowering the participation costs in the notice and comment process, these results reinforce previous observations that interest-group motivated mass comments now represent almost all public comments submitted on a docket. Although difficult to discuss quantitatively, even most non-mass comments submitted on the Broadband Privacy Order docket by individuals were non-substantive in nature; most simply appealed to how Obama was overstepping his bounds, delved into solely anti-government rhetoric, or even occasionally pure racism. The plethora of repetitive and values-based comments act exactly as scholars like Shulman have previously observed, serving to crowd out the handful of substantive comments with hundreds of thousands of wholly irrelevant ones. It is no wonder, then, that policymakers tasked with reading comments wonder how many people truly care about the issue, and fail to even comment on the volume or sentiment of comments from individuals in the final rule.

From the sheer number of mass, non substantive comments submitted on the Broadband Privacy Order docket, it is also clear that the mass public does not understand how the public comment system works. As scholars like Shulman and Mendelson have observed, the sheer

number of values-based comments, often motivated by interest group messaging, point to a mass belief that policymakers can or do take into account the volume of comments taking a particular stance when making policy decisions; this, of course, is legally prohibited. As Mendelson (2011) points out at length, this misunderstanding only serves to foster additional distrust. In dockets like the Broadband Privacy Order, where most comments reflect a level of distrust or hatred towards government already, this misunderstanding will actively foster resentment and a belief that government does not listen in groups where the trust-building benefits of participatory democracy would prove most useful.

Additionally, there are clear demographic differences between the general population and those commenting on the Order, which would not necessarily be expected from publicly available polling. This is not inherently a problem, as the process was never intended to be majoritarian; these results also do not account for the combined potential for differential issue salience by demographic group, which may help explain the differences somewhat. However, these findings do indicate that e-rulemaking may not include a diverse array of individuals and thoughts in the process as originally intended. It additionally points to pre-existing concerns surrounding the digital divide; individuals without broadband access are less likely to comment on the Order, even controlling for demographic factors.

Combined, these results show that more participation does not necessarily mean more engagement on the issue, and additionally that there are distinct differences in participation by demographic group. Twenty years on, e-rulemaking is clearly not accomplishing its goals of inclusivity and substantive democratic engagement, and has certainly not helped the public understand the extremely impactful rulemaking process. It could even be argued that e-rulemaking is proving bad for democracy; those that participate do not feel like they are heard, breeding additional resentment and distrust in government. While this paper will not go that far, it is clear that federal agencies need to reconsider their approach to including individuals in the notice and comment process.

The addition of two sets of information would help individuals better understand the rulemaking process and submit more-useful comments: easily understandable documents about the rule from rulemaking agencies like the FCC, and an education campaign to inform commenters of what comments are intended to do in notice and comment rulemaking. The FCC and other agencies currently post dozens to hundreds of detailed pages of information and requests in their Notice of Proposed Rulemaking documents. While this is extremely helpful for interested organizations with highly-knowledgeable experts, agencies should also produce documents meant to be read by the general public. A short document (1 to 2 pages) with a brief summary of the issue as seen by the agency, the agency's main intended decision, and several key questions would go a long way towards helping the general public to understand what is actually being discussed. This might be a lot of effort for little direct measurable gain — and may be susceptible to manipulation by the agency — but is likely necessary to solicit meaningful public comment from individuals in the mass public as e-rulemaking originally intended. Additionally, a public information campaign to educate individuals on what a public comment is and is not intended to do would help alleviate popular misunderstandings. This does not need to be advertisements — although I welcome the attempt to make notice and comment rulemaking sound interesting to the mass public — but should include clear banners on the FCC's ECFS system (and equivalent at other agencies), pop-ups, and obviously noticeable help text. While this would not necessarily alleviate misunderstandings in mass comments from organizations that spoof the FCC's submit page right away, it would at least help educate individuals interested in finding out how to write an effective public comment.

Manufactured Salience

The results of this paper show that the vast majority of mass comments came from Protect Internet Freedom, which this paper additionally aims to credibly demonstrate was

an astroturfing organization involved in manufacturing salience in individuals. Individuals expressing their identity through words and mechanisms recommended by external organizations is not, in my opinion, inherently normatively problematic — after all, interest groups and political parties necessarily do this on a daily basis. The trouble comes when these organizations are involved in astroturfing — hiding their true motivations and funding source — in an attempt to create a fake grassroots movement. Whether they agree with the message or not regardless or origin, individuals ought to know who is behind an organization giving them a message; otherwise, they may be susceptible to manipulation.

Realistically, though, this problem of concealing funding sources and intent cannot be solved easily, if at all. As such, the FCC or other federal agencies should take actions to decrease the incentives for manufactured salience by limiting the number of form letters that can be sent as individual comments. A quick solution to this would be to collate comments by message similarity; similar to a petition, a set of many near-identical messages could be collapsed into one broad folder (or message), with commenter names and any added text displayed below. Thanks to modern computing power, these text-similarity computations can be done in minutes, reducing the workload for agency staff. While this would not completely remove the incentives of interest groups and organizations manufacturing salience alike to submit near-identical comments, it would likely push them towards suggesting commenters write more-original comments in order to regain the loss in (limited) volume-based impact. Though not perfect, combined with previous recommendations on agency-provided information and a clearer understanding of the rulemaking progress, this would be a step towards the e-rulemaking goal of increasing participatory democracy.

Dataset Ethics

Finally, I feel it is important to close with a brief discussion of the highly personal name and address data of commenters that the FCC makes publicly available. There are two

interesting ethical questions here: first is whether individual commenters are informed of a choice between posting and not posting at all, and second, whether there are secondary use concerns. The first is not a new question, and has in fact been discussed by the Government Accountability Office recently.[82] The Office found that, by and large, federal agencies including the FCC should more clearly describe all data that will be publicly available after posting, as well as that the inclusion of a name, address, and email address is not required to post. Ironically, the Protect Internet Freedom form did a good (perhaps even a better) job of notifying posters that these data would be publicly available than the FCC itself, although in both cases address and email data was not clearly marked as optional. Assuming that the FCC and other federal agencies do this effectively, posters can therefore make an informed choice about whether to include their PII or not in a comment.

However, it is quite likely that at least some of the 150,000 individual commenters who left their name and address on the docket did not consider potential secondary data use concerns. After all, these data are personally identifying information (PII); a quick Google search of any name and address of a commenter on the docket easily reveals the vast majority of their life. At the scale of a dataset containing all commenters over a 10 year period, this is a significant portion of the US population. With a name and address, third parties could match this PII data to other large datasets, or simply use similar methods to this paper to target ads or build a detailed political opinion profile for personal or organizational gain; datasets on issue areas like these would be highly effective at attracting individuals to manufacture salience for astroturfing organizations. As it stands, the only real limitation to this is the speed of the free Census geocoding service, which takes about a second to geocode 8 potential addresses. This tension between personal anonymity, identity verification, and open data access from the FCC are unresolvable, but these issues can only become more relevant as individual public comments become more and more frequent. For my part, I do not include a technical appendix of code at the end of this paper for these reasons; while I

[82] U. S. Government Accountability Office, "Federal Rulemaking."

do not think this paper will be widely circulated, there is no reason to give motivated parties an easy opportunity to create such a targeted dataset. As such, the code and the full dataset from this paper will only be available upon request.

Conclusion

Over the past two decades, the move towards e-rulemaking has been a disruptive force in public comment participation; rulemaking proceedings that once would have been a docket of detailed comments from interest groups are now filled with comments by participating individuals from the mass public. However, while participation has doubtlessly increased by orders of magnitude, the benefits that proponents claimed the move to e-rulemaking would bring — heightened public understanding and an enhancement of the democratic process — have arguably failed to materialize. This paper furthers the existing literature on individual participation in public comment by examining what sorts of individuals participate, how they participate, what organizations (if any) motivate them to participate, and finally whether these three questions differ by sentiment towards the rule.

Through a novel dataset created from the Broadband Privacy Order docket — the 4th most commented-on Federal Communications Commission rulemaking proceeding — this paper finds evidence for a wide set of differences by party, race, sex, age, income, and broadband access across sentiment, comment frequency, likelihood of mass comments, and when compared to the general population. While some of these results support previous findings from the literature, others differ. I also find significant evidence of manufactured salience in the comment dataset, demonstrating that 94% of comments sent by individual commenters were likely mass comments from Protect Internet Freedom. This very-likely-astroturfing organization appeared to successfully target Republicans in order to motivate an extremely large volume of nearly-identical unsubstantial comments against the Order.

Combined, these results provide additional evidence to literature that suggests

e-rulemaking has not lived up to its goals of inclusivity and substantive democratic engagement. Additionally, a fundamental misunderstanding on the part of the mass public about the role of comments in rulemaking — that it is majoritarian, rather than substantive, in nature — has led mass public commenters to resent the government they are attempting to participate in when their comments are inevitably not acted on. As this paper shows, the vast majority of commenters on the Order were already anti-government in nature, and the fundamental lack of recognition or reaction to these comments by the FCC did nothing to assuage this sentiment. While agencies like the FCC have no obligation to change their proposed policy decisions due to an influx of negative and insubstantial mass comments, they should also not misleadingly write off these commenters entirely.[83]

Policy Implications

E-rulemaking — and thus the subject of this paper — are not necessarily limited in scope to agencies following the Administrative Procedures Act in the US federal government; though a comparison of notice and comment rulemaking at the state level could be a thesis in itself, a short survey of the administrative procedures acts in California, New York, and Florida reveals a patchwork of commenting procedures. Although all three states surveyed allow for electronic comment submittal — either via email or through a form — none of the states listed allowed for individuals to browse the text of other submitted comments on a rule, as the FCC and Regulations.gov do, nor were the instructions nearly as friendly or clear.[84,85,86] The European Commission has also recently funded a study into the feasibility of implementing e-rulemaking in a similar manner to the US, creating the possibility that this process may expand into Europe in the near future.[87] Perhaps unsurprisingly given the popularity of the concept, scholars in the literature have not explicitly called for the end of

[83] Broadband Privacy Order, 205.
[84] Madole, "Administrative Law: How California Regulations Are Made."
[85] "Instructions for Submitting Public Comments."
[86] "FLRules FAQ - FAC, FAR, ERulemaking."
[87] "Promoting E-Rulemaking in the EU through Deliberative Procedures."

e-rulemaking, although Bryer (2013) has suggested that if the costs of educating individuals to provide meaningfully substantive comment prove too high, then perhaps e-rulemaking should be phased out.

To modify e-rulemaking in the longer term, federal agencies like the FCC should consider publishing additional easily-understandable documents about the rule for mass public consumption, as well as begin an education campaign to inform individuals of what comments are intended and not intended to do in notice and comment rulemaking. Finally, to reduce the incentives for organizations intending to manufacture salience through nonsubstantive mass comments at scale, agencies like the FCC should also consider collapsing form comments into a petition-like comment with individual names and minor additions below. With these policy changes, agencies like the FCC may be able to move closer to their original goals of inclusivity and substantive democratic engagement.

Works Cited

Alan Douglas. 2016. *FCC Electronic Comment Filing System*. https://www.fcc.gov/ecfs/filing/10626974830836.

American Community Survey. 2010. "Median Income in the Past 12 Months." US Census Bureau. https://data.census.gov/cedsci/table?q=median%20household%20income&tid=ACSST1Y2019.S1901&hidePreview=false.

"Blog - Protect Internet Freedom." 2016. *Protect Internet Freedom*. https://web.archive.org/web/20170216040529/https://www.protectinternetfreedom.com/blog/.

Bryer, Thomas A. 2013. "Public Participation in Regulatory Decision-Making." *Public Performance & Management Review* 37 (2): 263–79. https://doi.org/10.2753/PMR1530-9576370204.

Carlitz, Robert D, and Rosemary W Gunn. 2002. "Online Rulemaking: A Step Toward E-Governance." *Government Information Quarterly* 19 (4): 389–405. https://doi.org/10.1016/S0740-624X(02)00118-1.

"Datamade/Usaddress." 2021. datamade. https://github.com/datamade/usaddress.

"ECFS FAQ." n.d. *Federal Communications Commission*. Accessed March 19, 2021. https://www.fcc.gov/ecfs/help/ecfs.

"E-Government Act of 2002." 2002. https://www.congress.gov/107/plaws/publ347/PLAW-107publ347.pdf.

"Electronic Comment Filing System Public API." 2020. Federal Communications Commission. https://www.fcc.gov/ecfs/help/public_api.

Evelyn Remaley. 2020. "NTIA Data Reveal Shifts in Technology Use, Persistent Digital Divide." *National Telecommunications and Information Administration*. https://www.ntia.doc.gov/blog/2020/ntia-data-reveal-shifts-technology-use-persistent-digital-divide.

Farina, Cynthia, Paul Miller, Mary Newhart, Claire Cardie, Dan Cosley, and Rebecca Vernon. 2011. "Rulemaking in 140 Characters or Less: Social Networking and Public Participation in Rulemaking." *Cornell Law Faculty Publications*, January. https://scholarship.law.cornell.edu/facpub/174.

"FCC Connect." 2016. *FCC Connect*. https://web.archive.org/web/20180409184313/https://connectfcc.com/.

"FCC Fixed Broadband Deployment." n.d. *FCC Fixed Broadband Deployment*. Accessed March 19, 2021. https://broadbandmap.fcc.gov/.

"FCC Releases Proposed Rules to Protect Broadband Consumer Privacy." 2016. *Federal Communications Commission*. https://www.fcc.gov/document/fcc-releases-proposed-rules-protect-broadband-consumer-privacy.

Figueiredo, John M. de. 2005. "E-Rulemaking: Bringing Data to Theory at the Federal Communications Commission Thirty-Fifth Annual Administrative Law Issue - The Role of the Internet in Agency Decisionmaking." *Duke Law Journal* 55 (5): 969–94. https://heinonline.org/HOL/P?h=hein.journals/duklr55&i=983.

"FLRules FAQ - FAC, FAR, eRulemaking." 2010. *Florida Department of State*. https://www.flrules.org/Help/newHelp.asp.

Gangadharan, Seeta Peña. 2009. "Public Participation and Agency Discretion in Rulemaking at the Federal Communications Commission." *Journal of Communication Inquiry* 33 (4): 337–53. https://doi.org/10.1177/0196859909340348.

Garvey, Todd. n.d. "A Brief Overview of Rulemaking and Judicial Review," 20.

House Speaker Paul Ryan. 2017. "Providing For Consideration Of S.J. Res. 34, Providing For Congressional Disapproval Of A Rule Submitted By The Federal Communications Commission." Congressional Record. https://www.congress.gov/congressional-record/2017/3/28/house-section/article/h2478-1.

"Huffington Post Online Privacy Poll." 2017. YouGov. http://big.assets.huffingtonpost.com/tabs_HP_Online_Privacy_20170330.pdf.

Ihkas Riapos. 2016. *FCC Electronic Comment Filing System*. https://www.fcc.gov/ecfs/filing/10626965901767.

"Instructions for Submitting Public Comments." 2021. *New York Department of State, Division of Administrative Rules*. https://www.dos.ny.gov/info/register/instruct.htm.

Jeff Kao. 2018. "Fighting Disinformation Using NLP." Minneanalytics Big Data Conference. https://github.com/j2kao/fcc_nn_research/blob/master/Presentation%20Kao%20MN%20Big%20Data%206-5-2018%20(final).pdf.

Jordan, Sara R., and K. Lee Watson. 2019. "Reexamining Rulemaking in an Era of Internet-Enabled Participation." *Public Performance & Management Review* 42 (4): 836–63. https://doi.org/10.1080/15309576.2018.1561377.

Kao, Jeff. 2021. "J2kao/Fcc_nn_research." https://github.com/j2kao/fcc_nn_research.

Larkin, Paul J. n.d. "Reawakening the Congressional Review Act." *Public Policy* 41 (1): 66.

Leighley, Jan E., and Jonathan Nagler. 2014. *Who Votes Now?: Demographics, Issues, Inequality, and Turnout in the United States*. Princeton University Press. http://www.jstor.org/stable/j.ctt4cgcqb.

Lois Grebosky. 2016. *FCC Electronic Comment Filing System*. https://www.fcc.gov/ecfs/filing/60001673422.

Madole, Amber. n.d. "Administrative Law: How California Regulations Are Made." Accessed April 11, 2021. https://guides.library.lls.edu/c.php?g=497706&p=6434545.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2020. "National Historical Geographic Information System: Version 15.0." Minneapolis, MN: IPUMS. https://doi.org/10.18128/D050.V15.0.

Mendelson, Nina A. 2011. "Rulemaking, Democracy, and Torrents of E-Mail." *George Washington Law Review* 79 (5): 1343–80. https://heinonline.org/HOL/P?h=hein.journals/gwlr79&i=1356.

———. 2012. "Should Mass Comments Count Response Essay." *Michigan Journal of Environmental & Administrative Law* 2 (1): 173–84. https://heinonline.org/HOL/P?h=hein.journals/michjo2&i=173.

"Mggg/Ohio-Precincts." 2021. Metric Geometry; Gerrymandering Group. https://github.com/mggg/ohio-precincts.

"Mggg-States/IN-Shapefiles." 2021. MGGG States. https://github.com/mggg-states/IN-shapefiles.

MIT Election Data and Science Lab. 2019. "U.S. President Precinct-Level Returns 2016," January. https://doi.org/10.7910/DVN/LYWX3D.

O'Connell, Anne Joseph. 2011. "Agency Rulemaking and Political Transitions." *NORTHWESTERN UNIVERSITY LAW REVIEW*, 64.

Olsen, V. Beth Kuser, Gerald E. Galloway, and Matthias Ruth. 2018. "The Demographics of Public Participation Access When Communicating Environmental Risk." *Human Ecology Review*, September, 115–36. https://doi.org/10.22459/HER.24.01.2018.06.

Phillip Dampier. 2009. "Broadband Usage Caps: 'Just Switch Providers' — George 'Out of Touch With Reality' Ou Misinforms (Again)." *Stop the Cap!* https://stopthecap.com/2009/08/26/broadband-usage-caps-just-switch-providers-george-out-of-touch-with-reality-ou-misinforms-again/.

"Promoting E-Rulemaking in the EU Through Deliberative Procedures." 2020. European Commission Community Research; Development Information Service. https://cordis.europa.eu/project/id/798502/reporting.

"Protect Internet Freedom Foundation." 2021. *IRS Tax Exempt Organization Search.*

https://apps.irs.gov/app/eos/detailsPage?ein=474548494&name=Protect%20Internet%20Freedom%20Foundation&city=Alexandria&state=VA&countryAbbr=US&dba=&type=CHARITIES,%20DETERMINATIONLETTERS,%20EPOSTCARD&orgTags=CHARITIES&orgTags=DETERMINATIONLETTERS&orgTags=EPOSTCARD.

"Protecting the Privacy of Customers of Broadband and Other Telecommunications Services." 2016a. Federal Communications Commission. https://ecfsapi.fcc.gov/file/11022064030581/FCC-16-148A1.pdf.

———. 2016b. *Federal Register*. https://www.federalregister.gov/documents/2016/12/02/2016-28006/protecting-the-privacy-of-customers-of-broadband-and-other-telecommunications-services.

"Public Notice and Comment Rulemaking (United States)." 2016. Organisation for Economic Co-operation; Development. https://www.oecd.org/gov/regulatory-policy/USA-Public-Notice-and-Comment.pdf.

rarohla. 2017. "2016 New York Primaries by Precinct." *CARTO*. https://rarohla.carto.com/tables/table_2016_ny_primaries_by_precinct/public.

Ryan Singel. 2018. "Filtering Out the Bots: What Americans Actually Told the FCC about Net Neutrality Repeal." The Center for Internet; Society. https://cyberlaw.stanford.edu/files/blogs/FilteringOutTheBotsUnique2017NetNeutralityComments1024Update.pdf.

"S.J.Res.34 - 115th Congress (2017-2018)." 2017. Webpage. https://www.congress.gov/bill/115th-congress/senate-joint-resolution/34/all-actions.

Schlosberg, David, and John S. Dryzek. 2002. "Digital Democracy: Authentic or Virtual?" *Organization & Environment* 15 (3): 332–35. https://doi.org/10.1177/1086026602153011.

Schlosberg, David, Stephen Zavestoski, and Stuart W. Shulman. 2008. "Democracy and E-Rulemaking: Web-Based Technologies, Participation, and the Potential for Deliberation."

Journal of Information Technology & Politics 4 (1): 37–55. https://doi.org/10.1300/J516v04n01_04.

Schlozman, Kay Lehman, Sidney Verba, and Henry E. Brady. 2010. "Weapon of the Strong? Participatory Inequality and the Internet." *Perspectives on Politics* 8 (2): 487–509. https://www.jstor.org/stable/25698614.

School, Stanford Law. n.d. "Rethinking Regulatory Democracy." *Stanford Law School*. Accessed March 19, 2021. https://law.stanford.edu/publications/rethinking-regulatory-democracy/.

Shulman, Stuart W. 2009. "The Case Against Mass E-Mails: Perverse Incentives and Low Quality Public Participation in U.S. Federal Rulemaking." *Policy & Internet* 1 (1): 23–53. https://doi.org/https://doi.org/10.2202/1944-2866.1010.

"Take Action - Free Our Internet." 2017. *Free Our Internet*. https://web.archive.org/web/20170516005556/http://freeourinternet.org/take-action/.

"Tell the FCC: Let Me Have Some Privacy." 2016. *Protect Internet Freedom*. https://web.archive.org/web/20171007222130/https://www.protectinternetfreedom.com/privacy-farce/.

"Tell the FCC's Chairman Wheeler: Your Privacy Rules Are a Farce!" 2016. *Protect Internet Freedom*. https://web.archive.org/web/20170216040343/https://www.protectinternetfreedom.com/projects/tell-the-fccs-chairman-wheeler/.

"The Federal Communications Commission (FCC)." 2021. *National Telecommunications and Information Administration*. https://www.ntia.doc.gov/book-page/federal-communications-commission-fcc.

"U.S. Census Bureau QuickFacts: United States." n.d. Accessed March 19, 2021. https://www.census.gov/quickfacts/fact/table/US/PST045219#.

U. S. Government Accountability Office. 2019. "Federal Rulemaking: Selected Agen-

cies Should Clearly Communicate Practices Associated with Identity Information in the Public Comment Process." GEO-19-483. Government Accountability Office. https://www.gao.gov/products/gao-19-483.

———. n.d. "Congressional Review Act." Accessed March 19, 2021. https://www.gao.gov/legal/other-legal-work/congressional-review-act.

"UPDATE: Throttled! Why Is the FCC Throttling and Blocking More Than 2,200 Opposition Comments from PIF Members?" 2016. *Protect Internet Freedom*. https://web.archive.org/web/20170216051035/https://www.protectinternetfreedom.com/2016/05/update-throttled-fcc-throttling-blocking-2200-opposition-comments-pif-members/.

US Census Bureau. n.d. "Median Age Doesn't Tell the Whole Story." *The United States Census Bureau*. Accessed March 19, 2021. https://www.census.gov/library/stories/2019/06/median-age-does-not-tell-the-whole-story.html.

"USA Census Block Group Boundaries, 2010." 2012. US Census Bureau. https://www.arcgis.com/home/item.html?id=1c924a53319a491ab43d5cb1d55d8561.

Voting and Election Science Team. 2021. "2016 Precinct-Level Election Results," March. https://doi.org/10.7910/DVN/NH5S2I.

"West Virginia Voting Districts from 2010 Census." 2012. Digital Maps; Geospatial Data, Princeton University. https://maps.princeton.edu/catalog/princeton-f1881q39x.

"What We Do." 2010. *Federal Communications Commission*. https://www.fcc.gov/about-fcc/what-we-do.

Appendix 1:

Set of All Unique Comments and Classified Sentiment

Warning:

These Unaltered Comments May Contain Offensive and Racist Language.

See Them Here:

https://docs.google.com/spreadsheets/d/1LUZZRA1ogjeVUBhi2l2ad3fi6GukiPBX_s6kRpzHDA8/edit?usp=sharing

Appendix 2:

Details of Block Group Level Prediction Reliability Testing

See *Reliability of Block Group Level Data* for context

Comparisons Between Voter File Data and Block Group Predictions

	Florida		**North Carolina**	
	Voter File Probabilities	Block Group Predictions	Voter File Probabilities	Block Group Predictions
	(1)	(2)	(3)	(4)
White Variable	0.00016238***	0.00036052***	0.00014015***	0.00011170***
	(0.00000870)	(0.00003408)	(0.00001245)	(0.00004189)
Male Variable	0.00013004***	0.00044791***	0.00012168***	0.00029422***
	(0.00000830)	(0.00008039)	(0.00001182)	(0.00010793)
Age Variable	0.00000460***	0.00000337***	0.00000351***	0.00000468***
	(0.00000022)	(0.00000065)	(0.00000030)	(0.00000138)

Note: ***p<0.01