Nudging Civility in Online Discourse: An Experimental Approach

A thesis submitted by

Jing Liu

in partial fulfillment of the requirements for the degree of

Master of Science

in

Economics

Tufts University

May 2025

Advisor: Professor Laura Gee

Abstract

This study investigates whether behavioral nudges can promote civility in online discourse. Using an online experiment with 122 U.S. participants, we test the effects of two interventions: a civility oath and a public visibility condition, where participants are informed their responses may be posted online. Participants responded to politically charged statements with pre-generated response options varying in tone and content. The primary outcome measures whether a participant selected a civil response. Results from both OLS and probit regressions show that neither the civility oath nor the public visibility condition significantly increased the likelihood of civil behavior. While the interaction of the two treatments showed a small positive effect, it was not statistically significant. These findings suggest that simple nudges may be insufficient to improve civility in online environments and that more robust interventions may be necessary to foster constructive digital engagement.

Contents

List of Tables

1 Introduction

Civility is a cornerstone of democratic discourse, enabling productive dialogue and fostering mutual respect in public discussions. It plays a crucial role in maintaining social cohesion, reducing polarization, and facilitating meaningful political engagement (Herbst, 2010). However, while civility has long been considered an essential ideal in public life, maintaining it has always been a challenge. The rapid expansion of digital communication, particularly on social media, has intensified this challenge, making online incivility a growing concern.

Social media platforms provide spaces for political and social discussions, but they have also amplified incivility, including personal attacks, misinformation, and inflammatory rhetoric (Sobieraj and Berry, 2011). Unlike face-to-face conversations, where social norms and immediate social consequences shape behavior, online discussions lack many of these constraints, allowing uncivil interactions to spread quickly and with little accountability (Kiesler et al., 1984). The deterioration of online discourse is widely recognized as a problem, and many Americans view civility as essential to democratic functioning. For instance, a 2023 poll by the Georgetown Institute of Politics and Public Service found that 94% of voters agree that "respect for each other is the first step in having a government that works" (Georgetown Institute of Politics and Public Service, 2023). Furthermore, incivility has been linked to negative consequences, such as increased political disengagement, distrust in institutions, and heightened social division (Gervais, 2015).

Given these concerns, this study examines whether subtle interventions can promote civility in online discussions. Specifically, we investigate the effects of two behavioral nudges: (1) a civility oath, where participants pledge to engage respectfully before responding to online content, and (2) a public visibility treatment, where participants are informed that their responses may be publicly viewable. These interventions are based on the idea that individuals may regulate their behavior when they make explicit commitments or when they anticipate public scrutiny (Barberá, 2015; Bond et al., 2012).

To test these hypotheses, we designed an online experiment in which participants responded to real statements from online discussions, sourced from *Politifact*[1]. Each participant was randomly assigned to one of four experimental conditions that varied based on whether they took a civility oath and whether their responses were publicly visible. The key outcome of interest is whether participants selected a civil response from a set of pre-generated options. Additionally, the study collects data on participants' perceptions of truthfulness, political orientation, and demographic characteristics. While these factors are not analyzed in this paper, they could be valuable for future research.

The results of this study provide insight into whether simple, low-cost interventions can encourage more respectful interactions in digital spaces. Understanding what drives online incivility and how to mitigate it is particularly important in an era where digital discourse shapes public opinion and political engagement. The remainder of this thesis is structured as follows: Section 2 discusses how this study expands existing literature; Section 3 describes the experimental design; Section 4 discussed the data; Section 5 outlines the empirical strategy; Section 6 presents the results, including robustness checks and subgroup analyses; and Section 7 discusses broader implications for digital discourse and potential policy interventions.

2 Literature Review

In an era of increasing political polarization and global tensions, restoring civility in public discourse has become a pressing policy challenge. Theoretical literature suggests that a critical mass of polite users is needed for civility to spread within a network; however, if incivility dominates, it can create a self-reinforcing cycle of negative behavior (Antoci et al., 2016).

To address abusive behavior and misinformation, policymakers have largely taken a reg-

[1]PolitiFact is a fact-checking website that rates the accuracy of claims by elected officials and others on its Truth-O-Meter.

ulatory approach. Several countries have enacted laws aimed at reducing harassment and cyberbullying, such as the United Kingdom's *Malicious Communications Act*. Governments have also pressured social media platforms to adopt content moderation policies, leading companies like X and Instagram to deploy AI and human moderators to detect and remove hate speech, harassment, and other forms of incivility. These platforms have also published community guidelines outlining acceptable behavior and provided users with tools to report abusive content (Instagram Help Center, n.d.; X Help Center, n.d.).

Beyond government action, non-profit organizations have sought to promote civility through public awareness efforts, digital literacy campaigns, and online civility pledges. One example is Microsoft's *Council for Digital Good*, which encourages young people to take the *Digital Civility Challenge*, a pledge that includes commitments such as "We pledge to be kind; Look from others' perspectives." However, the effectiveness of these nudges remains uncertain. Even if civility appears to improve following such initiatives, attributing the change solely to the pledge is challenging due to potential confounding factors, such as changes in platform policies. Moreover, it is difficult to separate the effect of pledging from the impact of making online interactions publicly observable. This study aims to experimentally test how two behavioral nudges influence civility in online settings.

This study contributes to the growing experimental literature on the effects of pledges across various settings. Prior research shows that oaths reduce dishonesty in both individual decision-making (Zickfeld et al., 2023) and strategic interactions (Beck, 2021; Jacquemet et al., 2024), with public observability further discouraging dishonest behavior (Kingsuwankul and Villeval, 2023). To our knowledge, this is the first study to examine whether oath-taking influences civility in online discourse. We aim to disentangle two key mechanisms: the intrinsic motivation to uphold a pledge and the effect of public observability, the latter of which aligns with broader findings on social image concerns shaping prosocial behavior (Andreoni and Bernheim, 2009; Lacetera and Macis, 2010).

A key feature of our design is the mandatory nature of oath-taking for participants in

the treatment group, eliminating self-selection effects. Prior studies suggest that individuals who voluntarily take oaths may be more likely to adhere to them due to personal preferences for promise-keeping (Kingsuwankul and Villeval, 2023). By removing this element of choice, our study isolates the causal effect of oath-taking on online civility. This has practical implications for social media platforms considering pledges as a moderation tool.

Beyond oath-taking, our study contributes to the broader literature on online behavior. While much existing research focuses on misinformation—demonstrating that false information spreads faster than the truth (Vosoughi et al., 2018) and that inattention, rather than political bias, drives its sharing (Pennycook et al., 2020)—our study shifts attention to the tone of online responses, regardless of content accuracy. We also explore how oath-taking interacts with political ideology and gender, extending research such as Afrouzi, Arteaga, and Weisburst (Afrouzi et al., 2024), which finds that message content and the identity of its source shape belief formation. These insights informed the design of our survey, particularly the section assessing participants' perceptions of statement authorship and credibility.

Finally, our study contributes to research on belief formation and its influence on behavior (Armona et al., 2019; Bachmann et al., 2015; Bordalo et al., 2020; Giglio et al., 2021; Kuchler and Zafar, 2019; Malmendier and Nagel, 2011; Roth and Wohlfart, 2020). Prior work demonstrates that beliefs, shaped by experiences and biases, impact decision-making and social interactions. We extend this literature by examining how perceptions of political and ideological alignment affect online civility. Specifically, we test whether disagreement with a statement increases incivility, contributing to research on motivated reasoning and affective polarization. Additionally, we analyze whether shared political ideology fosters greater civility or exacerbates in-group policing. Similarly, we investigate whether gender dynamics influence civility, particularly in discussions on gender-related topics. Furthermore, we assess whether knowing the author of a statement affects participants' civility levels. These findings contribute to the broader discourse on political communication and digital behavior, offering insights into how social identity and belief alignment shape interactions in online spaces.

3 Experiment

All participants signed a consent form before beginning the survey. Initially, they were provided with an overview of the survey and what to expect. Participants were then randomly assigned to one of two treatments, resulting in four experimental groups, as shown in Table 1.

Following the treatments, participants proceeded to the main survey, in which they selected their potential responses to specific statements as if they had encountered them online. They also provided their perceptions of each statement. To ensure data quality, attention check questions were embedded within the survey. After choosing their responses, participants were also asked about their perceptions of each statement. Finally, participants completed demographic questions.

3.1 Private vs. Broadcast Response

At the start of the survey, participants were presented with an information screen explaining the general structure of the study. They were informed that they would see a series of posts (hereafter referred to as *statements*) sourced from actual online discussions. For each statement, they had to select a response from a pre-specified list of four options that best reflected how they would react in a public online forum.

Participants assigned to the Broadcast Response treatment (BNo, BO) received an additional message displayed in a bordered box, informing them that a randomly selected subset of responses would be anonymously posted in the comment sections of the original statements. This message highlighted that their responses might be publicly visible to anyone reading the comments online. A designated account was used to post these responses after the experimental sessions concluded.

To ensure that all participants understood their assigned condition, they had to answer a multiple-choice comprehension check after reading the instructions, shown below:

Comprehension Check

According to the instructions provided above, is it possible that your responses could be made publicly visible?

- ❍ *Yes, my responses could be made publicly visible, but they would only be posted anonymously without any identifying information.*
- ❍ *No, my responses will not be made publicly visible.*

If a participant selected an incorrect response—for example, if a participant in the BO or BNo group chose the second option—a pop-up message appeared with the text "Not Correct!" Subjects had to acknowledge this message before proceeding.

This feature ensured that participants fully engaged with their treatment condition rather than simply skimming the instructions, reducing the likelihood of misunderstandings about whether their responses could have been publicly visible.

3.2 Oath vs. No Oath

Participants assigned to the Oath treatment (PO, BO) were required to read and type a pledge that promoted civil online behavior. This message emphasized respectful engagement and discouraged offensive language or personal attacks in online discussions. To ensure participants actively engaged with the content, they had to accurately type the pledge to proceed. Successfully completing this task earned them a $1 bonus, serving as a nudge to encourage civility in digital interactions. Additionally, a minimum time requirement of 30 seconds was enforced on this page to prevent participants from copying and pasting the text without reading it.

In contrast, participants in the No-Oath treatment (PNo, BNo) were asked to type a neutral passage about the internet. This passage had a similar word count and structure to the pledge but did not include any explicit encouragement of civil behavior. The purpose

of this condition was to control for the effects of the typing task itself, ensuring that any observed differences between groups stemmed from the content of the pledge rather than the act of typing. Like the Oath treatment, participants in the Control condition had to accurately type the passage to proceed and received a $1 bonus for successful completion.

The specific text for each condition is shown below:

Typing Task

Oath Treatment: Participants assigned to the oath condition (PO, BO) were required to type the following pledge:

> *"I pledge to promote civility and respect in all my online interactions. I will not use offensive language or attack others based on their race, gender, or socioeconomic level. I will strive to engage in constructive dialogue, listening with empathy, and responding with kindness."*

No Oath Treatment: Participants assigned to the control condition (PNo, BNo) were required to type the following passage:

> *"The internet began evolving in the 1960s and became publicly available in the early 1990s. It has transformed global communication and introduced new considerations such as privacy, reflecting its complex role in society."*

3.3 Statements

The statements used in the survey were sourced from *Politifact* and were verified to exist in public online forums where users had actively engaged in discussions. These statements covered a range of topics, including politics, crime, immigration, and the COVID-19 pandemic.

The response options were generated using *ChatGPT-4o*, the most advanced model available at the time of the study's design and IRB approval. Each statement was categorized

based on its broad topic and its truthfulness classification (true, false, or half-true) according to *Politifact*'s fact-checking platform. The responses were produced using the following prompt:

> *"Please provide potential agreeing and disagreeing reactions to the statement in quotes, varying in vehemence and civility, based on typical social media comments and replies on similar posts."*

For example, consider the statement:

> *"Joe Biden is letting millions of people from jails, from prisons, from insane asylums, from mental institutions, drug dealers pour in."*[2]

Although this statement served as an illustration, it was not included in the actual experiment. *ChatGPT* generated several agreeing and disagreeing responses, varying in *civility*. The four selected responses for this example are shown in Table 2.

Uncivil responses were characterized by capitalization, excessive exclamation marks, and an overall hostile tone. Some responses included hashtags to enhance the ecological validity of an online forum setting. However, to ensure ethical compliance, no extremely offensive or abusive language was used. All statements and responses were reviewed and approved by the IRB.

Using this method, we compiled a pool of 16 statements. [3] Each participant responded to a total of seven statements—six serving as primary experimental statements and one as an attention check. The statements were presented in a random order, with each appearing on a separate screen. Additionally, the response options were randomized at the participant level to control for order effects. We have a number of other design elements which are beyond the scope of this paper's discussion. We discuss those in the Appendix.

[2]Source: Donald Trump; stated on May 29, 2024, in comments to reporters. Politifact
[3]See Appendix for all 16 statements.

3.4 Attention Check

High-quality survey and experimental research required participants to carefully read and engage with the questions and treatments. However, inattention could lead to inconsistent responses, reducing data reliability. To ensure data quality, we incorporated five attention check questions throughout the survey.

The first attention check appeared in the middle of the six statements each participant saw. It mirrored the format of all other statements, presenting four response options. However, it included an additional instruction—highlighted in red—explicitly directing participants to select the second option.

The second to fifth attention checks were embedded within the statement perception questions:

- **Truthfulness Assessment:** Participants were instructed to select *Half-True/Half-False* regardless of their actual opinion.
- **Political Orientation & Gender Attribution:** Participants had to move the cursor to a value between 60 and 90 on a 0-100 scale.
- **Author Identification:** Participants were required to type *"Bryan Steil"*.

In each case, the special instructions were prominently highlighted in red to ensure visibility. These checks helped verify participant attentiveness, improving the validity of the collected data.

3.5 Demographic Questions

To get a clearer picture of who our participants were and to explore how different backgrounds might have influenced responses, we included a set of demographic questions. These covered basic personal details such as gender, age, race/ethnicity, education, employment status, political views, and religious affiliation.

Participants could select their gender from several options, including non-binary and an open field for those who preferred to specify their own. For race and ethnicity, they could check multiple boxes to reflect their identity, with a separate question specifically asking whether they identified as Hispanic/Latino/a/x. We also asked about education, ranging from less than high school to doctoral and professional degrees, and included a question about employment status with options like full-time, part-time, unemployed, student, or retired. To get a sense of political ideology, we used a five-point scale from "Very Liberal" to "Very Conservative," with options for "Don't Know" and "Prefer not to answer." Religious affiliation was measured with a multiple-choice question that listed major world religions, along with "None" and an open-ended "Other" option. These questions helped us understand the diversity of our participant pool and allowed for deeper analysis of how different demographic groups engaged with the survey.

4 Data

The survey was designed using Qualtrics and administered via Prolific, targeting U.S. participants. A total of 148 participants completed the survey, of whom 122 passed the attention check. To ensure data quality and validity, this paper focuses on the final sample of 122 participants. The sample was distributed across four experimental conditions: 31 participants in the Private-NoOath group, 29 in the Broadcast-NoOath group, 32 in the Private-Oath group, and 30 in the Broadcast-Oath group. The sample exhibited demographic diversity in terms of age, gender, education level, and political affiliation. A full summary of demographic characteristics is provided in Table 3. To ensure the success of randomization, Table 4 presents balance tests across treatment groups, showing no significant differences in most demographics.

In the survey, participants were presented with a series of statements and asked to choose a response from four available options, two of which were classified as civil. The main

dependent variable, *civil*, was a binary indicator that equaled 1 if a participant selected a civil response and 0 otherwise. The primary independent variables included *Broadcast*, a binary variable indicating whether the participant was informed that their responses might be publicly visible; *Oath*, a binary variable indicating whether the participant received the civility oath treatment; and *BroadcastOath*, an interaction term capturing the combined effect of both treatments. The proportion of civil responses in each experimental group is summarized in Table 2.

Beyond treatment effects, the survey also captured participants' perceptions of each statement, including agreement level, perceived political leaning, and author identification. These subjective measures helped assess how different ideological perspectives influenced response behavior. Additionally, demographic information was collected to explore potential heterogeneity in treatment effects.

For the OLS regression, we excluded participants who selected "Prefer not to answer" for political and religious affiliation questions. Only one and three participants, respectively, chose this option, making the exclusion negligible but simplifying result interpretation. After this data-cleaning process, the sample included 118 participants in the OLS regression without demographic controls. However, since one participant skipped all demographic questions, the effective sample size decreased to 117 participants when including demographic controls. As each participant responded to six statements, the total number of observations was 708 in the OLS regression without controls and 702 in the OLS regression with controls.

The next section outlines the empirical strategy employed to analyze the effects of the civility oath and public visibility treatments.

5 Empirical Specification

To examine the factors influencing online civility, we estimated the following linear regression model:

$$\text{Civil}_{is} = \beta_0 + \beta_1 \text{Oath}_i + \beta_2 \text{Broadcast}_i + \beta_3 \text{BroadcastOath}_i + \Gamma \text{X}_i + \epsilon_{is} \quad (1)$$

where $Civil_{is}$ was a binary outcome variable indicating whether respondent i selected a civil response in the survey to statement s. Recall each person saw six statements. The key independent variables included $Oath_i$, an indicator for whether the respondent was assigned to the civility oath condition, and $Broadcast_i$, an indicator for whether their responses were publicly visible. The interaction term $BroadcastOath_i$ captured potential heterogeneous treatment effects when both conditions were present. The model also included a vector of control variables, X_i, which accounted for respondent characteristics such as gender, race, education, employment status, political ideology, and religious affiliation. The error term ϵ_{is} captured all unobserved individual and statement-specific influences that affected civility responses but were not directly modeled through the included variables.

Given the experimental design, where participants were randomly assigned to the *Oath* and *Broadcast* visibility treatments, the coefficients on these variables could be interpreted as causal effects under the assumption of successful randomization. However, the inclusion of demographic controls accounted for residual variation that might have arisen from systematic differences in baseline characteristics across individuals.

We estimated the model using ordinary least squares (OLS) with robust standard errors clustered at the respondent level, as each respondent provided six responses throughout the survey. Although *Civil* was a binary outcome, OLS remained an appropriate estimation method as it provided unbiased estimates and facilitated straightforward interpretation of marginal effects.

For categorical control variables, the omitted categories served as reference groups: *female/non-*

binary/other for gender, *White* for race, *less than high school* for education, *employed full-time* for employment status, *moderate political ideology* for political affiliation, and *non-religious* for religious affiliation. This allowed for a clear interpretation of the estimated coefficients, where each coefficient reflected the difference in the likelihood of selecting a civil response relative to the corresponding reference group.

6 Results

6.1 OLS Regression Results

Table 5 presents the results of the ordinary least squares (OLS) regression estimating the likelihood of selecting a civil response. The coefficients on the main treatment variables—*Oath*, *Public*, and their interaction term *PublicOath*—remain statistically insignificant, indicating that neither the civility oath nor the public visibility treatment significantly influenced the likelihood of choosing a civil response. The coefficient on *Oath* is -0.0153 ($p = 0.750$), while *Public* is -0.0191 ($p = 0.658$). The interaction term *PublicOath* is also not significant, with a coefficient of 0.0349 ($p = 0.572$), suggesting that combining the two treatments does not produce a statistically significant joint effect on civility.

6.2 Robustness Check

To assess the robustness of the OLS results, we conducted a probit regression analysis, which is more appropriate for binary dependent variables such as civility. The probit model was estimated both without controls and with controls, and the results are presented in Table 6.

The results from the *probit model without controls* show that none of the treatment variables were statistically significant predictors of choosing a civil response. The coefficients for *Oath* ($-0.048, p = 0.826$), *Public* ($-0.184, p = 0.324$), and the interaction term *Public* $\times$ *Oath* ($0.208, p = 0.475$) were all small in magnitude. The overall *Pseudo* R^2 was 0.003,

indicating that the treatment variables alone explained very little of the variation in civility.

When demographic controls were included, the treatment effects remained statistically insignificant, confirming that the OLS findings were not driven by omitted variable bias. The overall *Pseudo* R^2 was 0.0846. For comparison, the adjusted R^2 for the OLS model with controls (Table 5) was 0.050, suggesting that the demographic and other control variables collectively explained about 5% of the variation in civility responses.

Overall, the probit results aligned closely with the OLS findings, providing robust evidence that the treatment variables did not significantly influence civility, while demographic characteristics played a more substantial role. The consistency of these results across model specifications suggested that the main findings were not sensitive to model choice. Full regression tables for both probit models are included in the appendix.

7 Conclusion

This study explored whether simple behavioral nudges, a civility oath, public visibility or the combination of these nudges, could promote more respectful interactions in online discussions. Using an online experiment with 122 participants, we examined whether committing to civility or knowing that responses could be visible to others would lead participants to choose more civil replies in a simulated comment section.

The findings suggest that neither intervention significantly influenced behavior when considered separately. The OLS regression results indicate that participants who took the civility oath were actually slightly less likely to choose civil responses compared to those in the control group ($\beta = -0.002$, s.e.= 0.049). Similarly, participants in the public visibility condition were also less likely to choose the civil option ($\beta = -0.046$, s.e.= 0.043), suggesting that public exposure alone did not encourage civility.

Interestingly, the interaction term between public visibility and the oath is positive ($\beta = 0.069$, s.e. $= 0.067$), implying that when both interventions were applied together, par-

ticipants were slightly more likely to choose civil responses than in the individual treatment conditions. However, this effect remains small and is not statistically significant, indicating that the combined intervention did not produce a meaningful shift in behavior.

One possible explanation is that the baseline level of civility in our sample was already high, particularly in the control group, leaving little room for the interventions to increase civility further. Additionally, it is possible that taking an oath without intrinsic motivation backfired slightly, making participants feel constrained rather than genuinely committed to civility. Other factors such as ingrained communication habits, psychological tendencies, or the overall tone of online discourse may play a more substantial role in shaping behavior than these interventions.

This indicates that fostering online civility may require more than a simple reminder or the prospect of public scrutiny. Factors such as ingrained communication habits, psychological tendencies, or the broader tone of online discourse may play a stronger role in shaping behavior than the interventions tested in this study. These results contribute to the ongoing discussion on digital civility and behavioral interventions. While previous research has demonstrated that public accountability and personal commitment can influence behavior in offline settings (Andreoni and Bernheim, 2009; Jacquemet et al., 2024; Kingsuwankul and Villeval, 2023; Lacetera and Macis, 2010), replicating these effects in online discussions appears to be more challenging. The findings suggest that platforms and policymakers seeking to improve online discourse may need to implement more structured interventions, such as community moderation, incentives for civil engagement, or content curation strategies.

However, this study has several limitations. The participant pool, while diverse, may not fully represent the broader population of social media users, particularly as Prolific users may differ from typical online communities. Additionally, the controlled nature of the experiment does not fully capture the complexity of real-world online interactions, where factors like anonymity, group dynamics, and emotional intensity play a significant role. Although key demographic and ideological differences can be controlled for in our regression models, other

factors such as past experiences with online conflict or levels of trust in digital platforms may have influenced participant responses. Future research could explore interventions that more closely mimic real-world online environments, such as integrating experiments into actual social media platforms or creating interactive simulations. Alternative strategies, such as personalized messaging, financial incentives for constructive engagement, or penalties for negative interactions, could also be tested. Additionally, studies could examine whether the effectiveness of civility interventions varies depending on the platform, or emotional tone of the discussion.

Although this study did not find strong evidence that a civility oath or public visibility significantly influences online behavior, it highlights the complexities of fostering respectful discourse in digital spaces. Given the increasing role of online discussions in shaping public opinion and political engagement, further research is needed to develop and refine strategies for promoting constructive and meaningful interactions online.

References

Afrouzi, Hassan, Carolina Arteaga, and Emily Weisburst (2024) "Is It the Message or the Messenger? Examining Movement in Immigration Beliefs," *Journal of Political Economy Microeconomics*, 2 (2), 244–297, 10.1086/728365.

Andreoni, James and B. Douglas Bernheim (2009) "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 77 (5), 1607–1636, 10.3982/ECTA7384.

Antoci, Angelo, Alexia Delfino, Fabio Paglieri, Fabrizio Panebianco, and Fabio Sabatini (2016) "Civility vs. Incivility in Online Social Interactions: An Evolutionary Approach," *PLOS ONE*, 11 (11), e0164286, 10.1371/journal.pone.0164286.

Armona, Luis, Andreas Fuster, and Basit Zafar (2019) "Home Price Expectations and Behavior: Evidence from a Randomized Information Experiment," *The Review of Economic Studies*, 86 (4), 1371–1410, 10.1093/restud/rdy038.

Bachmann, Rüdiger, Tim Berg, and Eric R. Sims (2015) "Inflation Expectations and Readiness to Spend: Cross-Sectional Evidence," *American Economic Journal: Economic Policy*, 7 (1), 1–35, 10.1257/pol.20130292.

Barberá, Pablo (2015) "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S.," *American Political Science Association*, 10.2139/ssrn.2631420.

Beck, Tobias (2021) "How the honesty oath works: Quick, intuitive truth telling under

oath," *Journal of Behavioral and Experimental Economics*, 94, 101728, 10.1016/j.socec.2021.101728.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler (2012) "A 61-million-person experiment in social influence and political mobilization," *Nature*, 489, 295–298, 10.1038/nature11421.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020) "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110 (9), 2748–2782, 10.1257/aer.20181219.

Georgetown Institute of Politics and Public Service (2023) "Battleground Civility Poll: September 2023 Results," https://politics.georgetown.edu/2023/09/20/battleground-civility-press-release-september-2023/, Accessed: 2025-03-30.

Gervais, Bryan T. (2015) "Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-Based Experiment," *Journal of Information Technology & Politics*, 12 (2), 167–185, 10.1080/19331681.2014.997416.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Andreas Weber (2021) "Climate Change and Long-Run Discount Rates: Evidence from Real Estate," *The Review of Financial Studies*, 34 (8), 3527–3571, 10.1093/rfs/hhaa009.

Herbst, Susan (2010) *Rude Democracy: Civility and Incivility in American Politics*, Philadelphia, PA: Temple University Press.

Instagram Help Center (n.d.) "Community Guidelines," https://www.facebook.com/help/instagram/477434105621119/, Accessed: March 14, 2025.

Jacquemet, Nicolas, Stéphane Luchini, Jason F. Shogren, and Adam Zylbersztejn (2024) "Commitment to the truth creates trust in market exchange: Experimental evidence," *Games and Economic Behavior*, 148, 279–295, 10.1016/j.geb.2024.09.011.

Kiesler, Sara, Jane Siegel, and Timothy W. McGuire (1984) "Social Psychological Aspects of Computer-Mediated Communication," *American Psychologist*, 39 (10), 1123–1134, 10.1037/0003-066X.39.10.1123.

Kingsuwankul, Sirada and Marie Claire Villeval (2023) "Why Do Oaths Work? Image Concerns and Credibility in Promise Keeping," *SSRN Electronic Journal*, 10.2139/ssrn.4574459.

Kuchler, Theresa and Basit Zafar (2019) "Personal Experiences and Expectations about Aggregate Outcomes," *Journal of Finance*, 74 (5), 2491–2542, 10.1111/jofi.12819.

Lacetera, Nicola and Mario Macis (2010) "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme," *Journal of Economic Behavior & Organization*, 76 (2), 225–237, 10.1016/j.jebo.2010.08.007.

Malmendier, Ulrike and Stefan Nagel (2011) "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *The Quarterly Journal of Economics*, 126 (1), 373–416, 10.1093/qje/qjq004.

Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand (2020) "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings," *Management Science*, 66 (11), 4944–4957, 10.1287/mnsc.2019.3478.

Roth, Christopher and Johannes Wohlfart (2020) "How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?" *The Review of Economics and Statistics*, 102 (4), 731–748, 10.1162/rest_a_00854.

Sobieraj, Sarah and Jeffrey M. Berry (2011) "From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News," *Political Communication*, 28 (1), 19–41, 10.1080/10584609.2010.542360.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018) "The Spread of True and False News Online," *Science*, 359, 1146–1151, 10.1126/science.aap9559.

X Help Center (n.d.) "Rules and Policies," https://help.x.com/en/rules-and-policies, Accessed: March 14, 2025.

Zickfeld, Jan H., Karolina A. Ścigała, Alexander Weiss, John Michael, and Pavlos Mitkidis (2023) "Commitment to honesty oaths decreases dishonesty, but commitment to another individual does not affect dishonesty," *Communications Psychology*, 1 (1), 27, 10.1038/s44271-023-00028-7.

Appendix

7.1 Tables

Table 1: Overview of Treatment Groups

Private Response; No Oath (PNo)	Private Response; Oath (PO)
Broadcast Response; No Oath (BNo)	Broadcast Response; Oath (BO)

Table 2: ChatGPT-Generated Response Options by Agreement and Civility

Response	**Category**
I agree. Biden's policies are endangering our society by allowing too many dangerous people in.	Agreeing; Civil
BIDEN'S A DISASTER!!! He's opening doors to all kinds of scum. We're not safe anymore!!!	Agreeing; Uncivil
This claim is completely unfounded and dangerous. Biden is not letting criminals run wild.	Disagreeing; Civil
This is NONSENSE!!! Completely unfounded and dangerous!!! This is fear-mongering at its worst!	Disagreeing; Uncivil

Table 3: Demographic Breakdown by Treatment Group

	Private-NoOath (PNo)	Broadcast-NoOath (BNo)	Private-Oath (PO)	Broadcast-Oath (BO)	Full Sample	Prefer not to Answer	N
Categorical Variable							
Female	0.48	0.62	0.69	0.57	0.59	0	72
Hispanic	0.10	0.03	0.19	0.17	0.12	0	15
Age	35	45	35	35	35	0	121
Race						0	
White	0.74	0.62	0.72	0.60	0.67		82
Black or African American	0.16	0.28	0.16	0.27	0.21		26
Asian	0.00	0.03	0.03	0.07	0.03		4
Other	0.10	0.07	0.09	0.07	0.08		9
Education						0	
4-year college degree (BA/BS)	0.32	0.34	0.38	0.40	0.36		44
Some college	0.39	0.34	0.16	0.13	0.25		31
High school or less	0.10	0.10	0.19	0.17	0.14		17
Other	0.19	0.21	0.28	0.30	0.25		29
Employment						0	
Employed full-time	0.55	0.38	0.41	0.57	0.48		58
Employed part-time	0.16	0.34	0.38	0.27	0.29		35
A student	0.16	0.03	0.03	0.10	0.08		10
Other	0.13	0.24	0.19	0.07	0.16		18
Political						1	
Liberal	0.55	0.48	0.35	0.60	0.50		60
Moderate	0.32	0.24	0.26	0.07	0.22		27
Conservative	0.13	0.24	0.32	0.33	0.26		31
Other	0.00	0.03	0.06	0.00	0.02		3
Religious						3	
None	0.35	0.31	0.25	0.30	0.30		37
Catholic	0.16	0.31	0.25	0.27	0.25		30
Other Christian	0.19	0.28	0.19	0.23	0.22		27
Other	0.29	0.10	0.31	0.20	0.23		27
Continuous Variable							
Take_risk	5.42	6.10	6.42	6.20	6.03	0	121
Patience	7.13	7.52	7.61	7.13	7.35	0	121
Give_cause	7.29	8.17	7.71	7.13	7.57	0	121
Attentive Proportion	31/35	30/37	31/45	30/31	122/148		

Note: The variable "Female" is binary, with a value of 1 indicating participants who identified as "female,"

and a value of 0 encompassing those who identified as "male," "binary," "other," or "prefer not to answer." The proportion presented here reflects the percentage of participants in each treatment group who identified as female. Participants selected from age ranges such as "Under 30," "30-39," ..., "70 or above." These ranges were converted to their midpoint values (e.g., 30, 35, 45, ..., 70) for numerical analysis, and the median age was calculated and reported for each treatment group. The "Other" category in *race* includes participants who identified as "American Indian or Alaska Native," "Middle Eastern or North African," "Native Hawaiian or Other Pacific Islander," "Other," or "prefer not to answer." The category "High school and less" in *education* includes participants choosing "less than high school" and "high school or ged." The "Other" category in *education* includes participants with educational attainment levels of "2-year college degree (Associates)," "Master's degree (MA, MS)," "Doctoral degree (PhD)," "Professional degree (MD, JD, DDS, etc.)," or "prefer not to answer." The "Other" category in *employment* includes participants who identified as "retired," "a homemaker," "unemployed but looking for work," "unemployed and not looking for work," "disabled," "other," or "prefer not to answer." The term "Liberal" in *political* refers to participants who selected either "very liberal" or "somewhat liberal." The term "Conservative" in this context refers to participants who selected either "very conservative" or "somewhat conservative." The "Other" category in *religious* includes participants who chose "don't know" and "prefer not to answer." The "Other" category in *Religious* includes participants who identified as "Protestant," "Muslim," "Jewish," "Hindu," "Buddhist," "Sikh," "Other," or "prefer not to answer." For continuous variables, the mean values are reported. The column *Prefer not to Answer* reflects the number of participants who selected the option "Prefer not to answer." The survey includes a total of 122 observations; however, the sample size is 121 in this table due to one participant omitting all demographic questions.

Table 4: Balance Test: P-values from Regressions Predicting Group Assignment

Demographic	**Private-NoOath**	**Broadcast-NoOath**	**Private-Oath**	**Broadcast-Oath**
Male	0.431	0.724	0.167	0.397
Hispanic	0.074	0.256	0.306	0.203
Black or African American	0.193	0.124	0.194	0.401
Asian	**0.003**	0.684	0.990	0.444
Other Race	0.629	0.997	0.503	0.395
Some College	0.436	0.504	0.491	0.481
2-Year College Degree	0.387	0.426	0.260	0.355
4-Year College Degree	0.666	0.369	0.498	0.292
Master's Degree	0.307	0.778	0.969	0.295
Doctoral Degree	0.844	0.951	**0.002****	0.116
Professional Degree	0.866	0.950	**0.002****	0.116
Employed Part-Time	0.223	0.618	0.198	0.396
Retired	0.374	0.788	0.348	0.280
Homemaker	0.545	0.299	0.452	0.845
Student	0.254	0.367	0.627	0.815
Unemployed Looking	0.308	0.595	0.191	0.536
Unemployed Not Looking	0.082	0.303	0.255	0.556
Disabled	0.661	0.823	0.257	0.096
Liberal	0.392	0.186	0.274	0.416
Conservative	0.215	0.232	0.270	0.291
Prefer Not to Answer	**0.001****	0.858	0.473	0.376
Religious	0.795	0.846	0.747	0.473

Note: p-values less than 0.05 are bolded and ** indicate significant imbalance.

Table 5: OLS Regression Results

Variables	**Civil**	**Civil (No Controls)**
Oath	-0.015	-0.011
	(0.048)	(0.053)
Broadcast	-0.019	-0.046
	(0.043)	(0.048)
Broadcast × Oath	0.035	0.052
	(0.062)	(0.073)
Male	-0.032	
	(0.037)	
Hispanic	-0.009	
	(0.052)	
Race Category		
Black or African American	-0.108**	
	(0.051)	
Asian	-0.048	
	(0.079)	
Other Race	-0.058	
	(0.074)	
Education Category		
Some College	-0.077*	
	(0.042)	
2-Year College Degree	-0.063	
	(0.053)	
4-Year College Degree	-0.113**	
	(0.047)	
Master's Degree	-0.128**	
	(0.057)	
Doctoral Degree	0.025	
	(0.078)	
Professional Degree	-0.670***	
	(0.062)	
Employment Category		
Employed Part-Time	0.067	
	(0.038)	
Retired	0.121**	
	(0.059)	
Homemaker	0.028	
	(0.060)	
Student	-0.024	
	(0.059)	
Unemployed Looking	-0.064	
	(0.093)	
Unemployed Not Looking	-0.532	
	(0.320)	
Disabled	0.106	
	(0.902)	
Political Category		
Liberal	0.046	
	(0.051)	
Conservative	0.039	
	(0.065)	
Prefer not to answer	0.089	
	(0.125)	
Religious	-0.096	
	(0.042)	
Constant	0.951***	0.856***
	(0.064)	(0.031)
Observations	702	708
R-squared	0.084	0.002

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The baseline mean of the dependent variable (Civil) is 0.85 in the control group. Robust standard errors are reported in parentheses and clustered at the subject level. Observations from participants who selected "Prefer not to answer" were excluded for easier interpretation of the regression results. The regression without controls includes 118 participants. The OLS regression with demographic controls includes 702 observations (117 participants $\times$ 6 statements), while the regression without controls includes 708 observations (118 participants $\times$ 6 statements). The difference is due to one participant skipping all demographic questions, resulting in missing values when controls were added. *Civil* is a binary outcome variable equal to 1 if the participant selected a civil response to the statement and 0 otherwise. *Oath* is an indicator variable equal to 1 if the participant was assigned to take the civility oath before responding to statements. *Broadcast* is an indicator variable equal to 1 if the participant was in the public treatment condition, meaning their responses were potentially visible to others. *BroadcastOath* is an interaction term capturing the joint effect of the civility oath and the public treatment condition. *Male* is a binary variable equal to 1 if the participant identified as male and 0 if they identified as female, non-binary, or another gender identity. *Hispanic* is a binary variable equal to 1 if the participant identified as Hispanic/Latino/a/x and 0 otherwise. *Religious* is a binary variable equal to 1 if the participant identified with a religious affiliation (e.g., Protestant, Muslim, Jewish, Hindu, Buddhist, Sikh) and 0 if they selected "None." *Other Race* includes participants who identified as "American Indian or Alaska Native," "Middle Eastern or North African," "Native Hawaiian or Other Pacific Islander," or selected "Other" or "Prefer not to answer." *Liberal* includes participants who identified as "very liberal" or "somewhat liberal" on the political ideology scale. *Conservative* includes participants who identified as "very conservative" or "somewhat conservative" on the political ideology scale.

Table 6: Probit Regression Results

Variables	**Civil**	**Civil (No Controls)**
Oath	-0.078	-0.048
	(0.212)	(0.227)
Broadcast	-0.076	-0.184
	(0.189)	(0.193)
Broadcast × Oath	0.168	0.208
	(0.274)	(0.301)
Male	-0.137	
	(0.153)	
Hispanic	-0.045	
	(0.257)	
Race Category		
Black or African American	-0.416**	
	(0.190)	
Asian	-0.261	
	(0.327)	
Other Race	-0.257	
	(0.298)	
Education Category		
Some College	-0.480	
	(0.274)	
2-Year College Degree	-0.411	
	(0.337)	
4-Year College Degree	-0.678*	
	(0.354)	
Master's Degree	-0.733**	
	(0.301)	
Doctoral Degree		
	(Omitted)	
Professional Degree	-2.354***	
	(0.345)	
Employment Category		
Employed Part-Time	0.293*	
	(0.175)	
Retired	0.664	
	(0.419)	
Homemaker	0.134	
	(0.324)	
Student	-0.119	
	(0.220)	
Unemployed Looking	-0.457	
	(0.407)	
Unemployed Not Looking	-1.879**	
	(0.846)	
Disabled	0.490	
	(0.479)	
Political Category		
Liberal	0.209	
	(0.215)	
Conservative	0.115	
	(0.267)	
Prefer Not to Answer	0.417	
	(0.664)	
Religious	-0.035	
	(0.188)	
Constant	1.627***	1.061***
	(0.326)	(0.138)
Observations	696	708
Pseudo R-squared	0.084	0.003

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Robust standard errors in parentheses clustered at the subject level. There are 708 observations in the model without controls and 696 observations in the full model, as one participants was dropped due to perfect prediction issues in education categories. *Civil* is a binary outcome variable equal to 1 if the participant selected a civil response to the statement and 0 otherwise. *Oath* is an indicator variable equal to 1 if the participant was assigned to take the civility oath before responding to statements. *Broadcast* is an indicator variable equal to 1 if the participant was in the public treatment condition, meaning their responses were potentially visible to others. *BroadcastOath* is an interaction term capturing the joint effect of the civility oath and the public treatment condition. *Male* is a binary variable equal to 1 if the participant identified as male and 0 if they identified as female, non-binary, or another gender identity. *Hispanic* is a binary variable equal to 1 if the participant identified as Hispanic/Latino/a/x and 0 otherwise. *Religious* is a binary variable equal to 1 if the participant identified with a religious affiliation (e.g., Protestant, Muslim, Jewish, Hindu, Buddhist, Sikh) and 0 if they selected "None." *Other Race* includes participants who identified as "American Indian or Alaska Native," "Middle Eastern or North African," "Native Hawaiian or Other Pacific Islander," or selected "Other" or "Prefer not to answer." *Liberal* includes participants who identified as "very liberal" or "somewhat liberal" on the political ideology scale. *Conservative* includes participants who identified as "very conservative" or "somewhat conservative" on the political ideology scale. *High School or GED* is an indicator for participants whose highest level of education is a high school diploma or GED equivalent. *Some College*, *2-Year Degree*, *4-Year Degree*, and *Master's Degree* are indicator variables for progressively higher levels of education. *Doctoral Degree* and *Professional Degree* categories were omitted due to collinearity or perfect prediction. *Employed Part-Time*, *Retired*, *Homemaker*, *Student*, *Unemployed but Looking for Work*, and *Unemployed and Not Looking for Work* are employment categories relative to the omitted category of *Employed Full-Time*.

7.2 Statement Perception Assessment

After completing their initial responses to the statements, participants proceeded to the second part of the survey, where they answered four perception-based questions about each statement they had viewed. Each statement was followed by four separate questions, each appearing on its own page:

1. **Truthfulness Assessment**: Participants provided a quick judgment on whether the statement was *True*, *Half-True/Half-False*, or *False*.
2. **Political Orientation**: Participants used a slider ranging from 0 to 100 to indicate how liberal or conservative they perceived the statement to be. A value of 0 represented *Very Liberal*, 50 represented *Neutral*, and 100 represented *Very Conservative*.

3. **Gender Attribution**: Participants assessed the likelihood that the statement was made by someone who identified as male or female, using a slider where 0 represented *Male*, 50 represented *Don't Know*, and 100 represented *Female*. Additionally, a "Not Applicable" checkbox was provided for cases where gender attribution might not have been relevant.

4. **Author Identification**: Participants provided their best guess of who had made the statement.

Participants had the opportunity to earn a $0.50 bonus for correctly answering the *truthfulness* and *author identification* questions, but only if the statement was randomly selected as the "statement that counts."

To accommodate common misspellings in the author identification task, we used *ChatGPT-4o* to generate the six most common misspellings for both the first and last name of each author. Any combination of these misspellings was considered correct, ensuring participants were not penalized for minor typographical errors.[4] For example, if the correct answer was *Elon Musk*, responses such as *Elan Mask* were also accepted.

To minimize the risk of participants looking up answers online, several measures were implemented. Each question was assigned a strict 45-second time limit, after which the survey automatically advanced to the next question. This restriction ensured that participants made their assessments quickly, reducing the likelihood of external searches.

Additionally, page tracking code was embedded within Qualtrics to monitor whether a participant navigated away from the survey page. If a participant left the survey while answering a question, the corresponding statement was marked as ineligible for the bonus. This feature acted as a deterrent against seeking external information.

To reinforce compliance, the restriction on leaving the page was explicitly communicated in the initial instructions. Moreover, reminders were displayed on every question page to ensure participants remained aware of the rule throughout the survey. These precautions collectively aimed to preserve the integrity of participants' responses and maintain the validity of the study.

7.3 Survey Statements and Response Options

Below are the 16 statements used in the survey, along with the corresponding response options.

1. **The United States is an outlier, one of only about half a dozen countries, without any guarantee of paid leave for new parents and/or other health care needs.**

[4] See Appendix 8.4 for details on common misspellings.

- **Agree Uncivil:** This is why America sucks!!! We can spend trillions on war but can't take care of new parents. WAKE UP, PEOPLE!
- **Agree Civil:** It's about time people realized how behind we are. Other countries prioritize families and well-being, and we should too. Shameful but true.
- **Disagree Uncivil:** STOP WHINING! If you want paid leave, find a job that offers it. We don't need the government to babysit us!!!
- **Disagree Civil:** While paid parental leave sounds nice, we need to consider the economic impacts and how it would be implemented here.

2. **Donald Trump deported less, believe it or not, than Barack Obama even did.**
   - **Agree Uncivil:** FINALLY, people are realizing the truth!!! Obama was the real deporter-in-chief, NOT Trump. The media needs to stop lying!!!!
   - **Agree Civil:** Despite the rhetoric, Trump actually deported fewer people than Obama.
   - **Disagree Uncivil:** This is a JOKE!!! Trump built his campaign on anti-immigrant rhetoric and policies. Stop spreading lies! #TrumpLies
   - **Disagree Civil:** Deportation numbers alone don't capture the full impact of Trump's immigration policies. Obama's approach was different and more humane.
3. **For African American families in particular, the homeownership rate remains relatively unchanged since 1968, the year the Fair Housing Act was signed into law.**
   - **Agree Uncivil:** This is a total disgrace!!! Decades have passed and nothing has changed!! The system is rigged against African American families!!
   - **Agree Civil:** It's really disappointing that African American homeownership rates haven't improved since 1968. We need to address the systemic barriers still in place.
   - **Disagree Uncivil:** This is just more race-baiting!! African Americans have had plenty of opportunities, and it's their own fault if they're not buying homes!!!
   - **Disagree Civil:** I'm skeptical of this claim. There are many factors that affect homeownership rates, and it's not just about systemic issues.
4. **As many as 75% of school shootings resulted from a gun that was not secured.**
   - **Agree Uncivil:** This is INFURIATING. Irresponsible gun owners are literally putting our kids' LIVES at RISK. Lock up your damn guns.

- **Agree Civil:** It's alarming that so many school shootings involve unsecured guns. We need stricter laws to ensure firearms are properly secured.
- **Disagree Uncivil:** This statistic is probably exaggerated. More anti-gun PROPAGANDA to push for gun control. Stop blaming gun owners!!!!
- **Disagree Civil:** I'm skeptical about this claim. We should focus on addressing the root causes of violence rather than blaming gun owners.

5. **"Wage increases have exceeded" the cost of inflation.**
   - **Agree Uncivil:** Finally, some good news!!! If you're still complaining, you're just lazy and looking for excuses. Stop whining and get to work!!!!
   - **Agree Civil:** It's great to see wages finally rising faster than inflation. This should help many families struggling to make ends meet.
   - **Disagree Uncivil:** What a load of crap!!! My wages haven't kept up at all. These stats are total lies, and anyone who believes them is an idiot!!!!
   - **Disagree Civil:** I'm skeptical about this claim. Many people still feel like they're not keeping up with rising costs. Let's look at the details.
6. **Donald Trump "proposed when he was president" that "he wanted to raise the gas tax up to 25 cents."**
   - **Agree Uncivil:** Typical Trump move!!! Just another way to squeeze more $ out of ordinary people while pretending to care about infrastructure.
   - **Agree Civil:** Trump's proposal to increase the gas tax might have been aimed at addressing budgetary needs for infrastructure. It's a complex issue that requires thoughtful discussion.
   - **Disagree Uncivil:** Trump proposing a gas tax hike??? That's JUST INSANE. He clearly didn't think about how this would hurt people who are already struggling.
   - **Disagree Civil:** Increasing the gas tax by 25 cents might not have been the best approach, as it could disproportionately affect low-income individuals. There are other ways to fund infrastructure without burdening everyday drivers.
7. **Public schools are now as segregated by race and class as they were in the 1960s.**
   - **Agree Uncivil:** This is absolutely DISGUSTING!!! We're back in the dark ages with segregation!! Our education system is a complete failure!!!!

- **Agree Civil:** It's deeply troubling that our schools are still so segregated by race and class. We need to address this issue urgently.
- **Disagree Uncivil:** This is total BS!!!! Schools are NOT segregated like they were in the 1960s!! Stop spreading LIES to push a divisive agenda!!!
- **Disagree Civil:** I think this claim is exaggerated. While there are issues with inequality, comparing it to the 1960s is misleading and unproductive.

8. **Under President Joe Biden, "Black unemployment is the lowest in American history."**
   - **Agree Uncivil:** FINALLY, some real progress! Biden is actually getting things done for Black Americans! Anyone who disagrees is just blind to the FACTS!
   - **Agree Civil:** This is a significant achievement. Biden's policies are clearly making a positive impact on Black employment rates.
   - **Disagree Uncivil:** This is total BS! Those numbers are MANIPULATED! Biden hasn't done anything to help Black Americans, it's all a sham!
   - **Disagree Civil:** I'm skeptical of these claims. We need to look at the broader context and see if these numbers truly reflect lasting improvements.
9. **In the state of Florida, they decided middle school students will be taught that enslaved people benefited from slavery.**
   - **Agree Uncivil:** This is absolutely DISGUSTING. Florida is rewriting history to justify atrocities. This is pure EVIL. #ShameOnYou
   - **Agree Civil:** It's outrageous that Florida would promote such a harmful and false narrative. This is a disservice to education and truth.
   - **Disagree Uncivil:** This is total BS!!!! No one in their right mind believes that crap. Stop spreading LIES to make Florida look bad. #StopTheLies
   - **Disagree Civil:** I doubt this claim is accurate. We need to see the actual curriculum before jumping to conclusions about what's being taught.
10. **There is "clear scientific consensus" that "hormonal birth control makes you fat, doubles risk of depression and triples risk of suicide."**
    - **Agree Uncivil:** See!!! Birth control is dangerous! Doctors and Big Pharma are lying to us!!! It's time to wake up and stop poisoning ourselves!!!

- **Agree Civil:** This is really concerning. We need to take these risks seriously and ensure women are fully informed about the potential side effects of hormonal birth control.
- **Disagree Uncivil:** This is total garbage!!! There's no scientific consensus on this!! Stop spreading fear-mongering lies about birth control!!
- **Disagree Civil:** I find this hard to believe. We need to look at the actual studies and evidence before accepting such alarming claims about hormonal birth control.

11. **"Each year, more than $1 billion is stolen from" workers in New York state through wage theft.**
    - **Agree Uncivil:** This is a DISGRACE!! Employers are stealing BILLIONS from hardworking people!!! Lock them up and throw away the key!!!!
    - **Agree Civil:** It's shocking that wage theft is this rampant. We need stronger laws and enforcement to protect workers from being exploited.
    - **Disagree Uncivil:** This is exaggerated BS!!! NO WAY are employers stealing that much!! It's just anti-business propaganda!!!
    - **Disagree Civil:** These numbers seem inflated. We should verify the data before jumping to conclusions about the extent of wage theft.
12. **Every booster you take, you're more likely to get COVID as a result of it.**
    - **Agree Uncivil:** The more boosters you get, the more you mess up your immune system. It's a big pharma SCAM, and people are just sheep for following it!!!! #VaccineSafety
    - **Agree Civil:** I've read some articles that suggest there could be a correlation between multiple boosters and increased infection rates. It's something worth investigating further to ensure public health safety.
    - **Disagree Uncivil:** What a load of GARBAGE!!! People like you are the reason misinformation spreads so easily. Get your facts straight! #VaccinesSaveLives
    - **Disagree Civil:** Boosters are designed to help maintain immunity. There may be cases of infection post-vaccination, but the overall benefits are well-documented.
13. **The number of people in the U.S. illegally is upwards of 20, 25, maybe 30 million.**
    - **Agree Uncivil:** This is OUT OF CONTROL!! Our borders are a JOKE, and these numbers prove it. We need to crack down hard and fast, or we'll LOSE our COUNTRY!!!

- **Agree Civil:** If those numbers are accurate, it highlights a significant challenge that needs to be addressed through comprehensive immigration reform.
- **Disagree Uncivil:** This is NONSENSE!!! There's NO WAY those numbers are accurate. Stop spreading LIES and misinformation to create panic!!!!
- **Disagree Civil:** That estimate seems quite high compared to other reports. It's important to rely on verified data to have an informed discussion about immigration policy.

14. **Tariffs are a "direct, regressive tax on Americans" and President Joe Biden's new tariffs on Chinese goods will "hit every family."**
    - **Agree Uncivil:** EXACTLY. These tariffs are going to hurt American families the most! Biden's policies are just making life more EXPENSIVE for everyone. #BidenFail
    - **Agree Civil:** Tariffs do tend to raise prices for consumers, so it's important to consider their impact on everyday Americans. Hopefully, there will be measures to mitigate this effect.
    - **Disagree Uncivil:** This is just alarmist rhetoric!!! Tariffs are a NECESSARY strategy to deal with unfair trade practices. Complaining about them WITHOUT understanding the bigger picture is short-sighted. #SupportBiden
    - **Disagree Civil:** While tariffs can lead to higher prices, they are often used as a tool to protect domestic industries and address trade imbalances. It's a nuanced issue that requires a balanced approach.
15. **Black women in the U.S. are "three to four times more likely to die in connection with childbirth than other women."**
    - **Agree Uncivil:** This is an OUTRAGE!!! The healthcare system is CLEARLY biased and failing Black women. How can anyone look at these numbers and not see racism at play? This country is failing them, and it's SHAMEFUL.
    - **Agree Civil:** The disparities in maternal mortality for Black women are deeply concerning. It's a stark reminder that we must urgently address systemic inequality in healthcare to protect the lives of Black mothers and ensure equal care for all.
    - **Disagree Uncivil:** I'm so sick of hearing racism being blamed for EVERYTHING. It's a lazy excuse that ignores the real issues! Stop using race to manipulate and divide people—it's pathetic and destructive.

- **Disagree Civil:** While there are certainly disparities in healthcare, it's important to consider other factors that may contribute to these statistics. Let's work towards finding comprehensive solutions.

16. **Potato chips, KitKat bars, and Viagra are not taxed in Wisconsin because they are considered "essential."**

    - **Agree Uncivil:** Who doesn't love stuffing their face with chips and KitKats? Wisconsin knows we need junk food to survive—FORGET about actual health!!! And Viagra? Sure, let's make sure the state supports "FUN" in the bedroom.
    - **Agree Civil:** It seems that Wisconsin has a sense of humor when it comes to what's essential! Potato chips and KitKats might be seen as essential for maintaining a happy life, even if they're not the healthiest options.
    - **Disagree Uncivil:** It's absolutely INSANE that junk food and Viagra are considered essential while things like menstrual products get taxed!!! Who thought this was a good idea? It's OFFENSIVE, it's sexist, and it's out of touch with reality.
    - **Disagree Civil:** It's surprising that potato chips and candy bars are tax-free given the obesity epidemic. Maybe it's time to rethink our priorities and include genuinely essential items.

7.4 Accepted Misspellings for Author Names in Bonus Allocation

Below are the accepted misspellings for each author name corresponding to the 16 statements shown to participants:

1. **Lisa Subeck**: Liza, Lissa, Lysa, Leesa, Liesa, Subek, Subbeck, Subech, Suebeck, Suback
2. **Ron DeSantis**: Ran, Ronn, Roon, Ren, Rohn, De Santis, Desantes, Desanties, Desanti, DeSantez
3. **Tim Scott**: Tom, Timn, Tin, Timm, Tum, Scot, Scotte, Skott, Sctott, Scottt
4. **Kamala Harris**: Kamela, Kamla, Kammala, Camala, Kamalla, Harriss, Haris, Harries, Harys, Hariss
5. **Joe Biden**: Jo, Jow, Jhoe, Jooe, Jae, Bidden, Bidean, Bidon, Binden, Bideon
6. **Nikki Haley**: Niki, Nicki, Nikky, Nikie, Nicky, Haly, Hailey, Halley, Hayley, Halay
7. **Marco Rubio**: Maro, Marko, Margo, Marcio, Marcco, Rubo, Robio, Rubbio, Rubiio, Rubino
8. **Jared Polis**: Jarred, Jered, Jaredd, Jarrad, Jaird, Poles, Pollis, Plois, Polas, Poliz

9. **Bobby Scott**: Boby, Bobbye, Bobbie, Bobe, Bobbi, Scot, Scotte, Skott, Scoth, Scottt
10. **Donna Brazile**: Dona, Donnah, Donne, Donnna, Donnae, Brazil, Brazille, Brazel, Brazzile, Brizile
11. **Kamala Harris**: Kamela, Kamla, Kammala, Camala, Kamalla, Harriss, Haris, Harries, Harys, Hariss
12. **Elon Musk**: Elan, Eon, Elyn, Elin, Ellon, Muks, Musc, Musck, Muxk, Mask
13. **Linda Rosenthal**: Lynda, Lida, Lyndaah, Lindah, Londa, Rosenthall, Rosentahl, Rozenthal, Rosnethal, Rosentha
14. **Ron DeSantis**: Ran, Ronn, Roon, Ren, Rohn, De Santis, Desantes, Desanties, Desanti, DeSantez
15. **Kamala Harris**: Kamela, Kamla, Kammala, Camala, Kamalla, Harriss, Haris, Harries, Harys, Hariss
16. **Melissa Agard**: Mellisa, Malissa, Melisa, Mellissa, Melessa, Aggard, Agrad, Agaird, Agart, Agerd