Telematics and Informatics 44 (2019) 101263

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Perceived exposure to and avoidance of hate speech in various communication settings

Matthew Barnidge[a,*], Bumsoo Kim[b], Lindsey A. Sherrill[c], Žiga Luknar[d], Jiehua Zhang[a]

[a] The University of Alabama, United States
[b] The Hebrew University of Jerusalem, Israel
[c] University of North Alabama, United States
[d] European Law Institute, Austria

ARTICLE INFO

Keywords:
Hate speech
Social media
Online community
Mobile messaging apps
Face-to-face communication

ABSTRACT

Social media platforms have been accused of spreading hate speech. The goal of this study is to test the widespread belief that social media platforms have a high level of hate speech in the eyes of survey respondents. Secondarily, the study also tests the idea that encountering perceived hate speech is related to avoiding political talk. The study analyzes data from a two-wave online survey (N = 1493) conducted before and after the 2018 U.S. Midterm Elections, and it estimates perceived exposure to hate speech across multiple venues: face-to-face, social media, mobile messaging applications, and anonymous online message boards. Results show that (a) respondents report higher levels of hate speech on social media in comparison to face-to-face communication and (b) there is a positive relationship between perceived exposure to hate speech and avoidance of political talk. Results are discussed in light of public conversations about hate speech on social media.

1. Introduction

A recent Washington Post op-ed dubbed 2018 "the summer of hate speech," illustrating the salience of the topic in the United States (Downes, 2018). In recent years, a cultural debate about both the ethics and legality of limiting speech has raged, ranging from protests on college campuses to individual bans from social media platforms. Social media platforms, including Facebook and Twitter (Guiora & Park, 2017), as well as online anonymous message boards such as Reddit or 4Chan (Marwick & Lewis, 2015), have been accused by media organizations (Gopalan, 2018; Lima, 2018) and scholars (Guiora & Park, 2017; Schmidt & Wiegand, 2017) of exacerbating the spread of hate speech. In the wake of tragic events such as the Charlottesville, Virginia protests and the Pittsburgh synagogue shooting, public conversations have turned toward reporting and preventing hate speech on social media platforms (Carroll & Karpf, 2018), and what, if anything, can be done about it from a legal standpoint when it occurs (Peters, 2018).

The goal of this study is to test whether the widespread belief that social media platforms have a high level of hate speech is borne out by survey respondents. Secondarily, the study also tests the idea that encountering hate speech (as it is perceived by the respondent) is related to avoiding political talk. The study analyzes data from a two-wave online survey (N = 1493) conducted before and after the 2018 U.S. Midterm Elections, and it makes within-subjects comparisons of perceived exposure to hate speech across multiple venues: face-to-face, social media, mobile messaging applications, and anonymous online message boards. The study relies not on third-party observations of hate speech, but rather on the perceptions of social media users themselves (Costello, Hawdon,

* Corresponding author at: Box 870172, Tuscaloosa, AL 35487, United States.
E-mail address: mhbarnidge@ua.edu (M. Barnidge).

https://doi.org/10.1016/j.tele.2019.101263
Received 19 March 2019; Received in revised form 17 July 2019; Accepted 31 July 2019
Available online 06 August 2019
0736-5853/ © 2019 Elsevier Ltd. All rights reserved.

Ratliff, & Grantham, 2016; Cowan & Hodge, 1996; Leets, 2001). While there are certainly drawbacks to this approach, reliance on self-reported perceptions also has its strengths, including: (1) recognition of the fact that there is no broadly accepted legal definition of hate speech, which makes third-party observations difficult; (2) control of individual-level idiosyncrasies via within-subjects comparisons, assuming that individuals consistently over- or underestimate hate speech across communication settings; and (3) the value that studying perception adds to our understanding of the influence of communication on political behavior, regardless of whether a given individual over- or underestimates their exposure to hate speech.

2. Literature review

2.1. Hate speech

From a legislative perspective, hate speech is notoriously difficult to define. In the United States, the First Amendment prohibits any government interference with freedom of speech, thereby limiting legal provisions or remedies for punishing discriminatory speech against disadvantaged groups or along lines of race/ethnicity, religion, sexual orientation, and disability (Walker, 1994). An exception is defamation law, which seeks to balance protection of individuals' reputations and the right to free speech (Weaver, Kenyon, Partlett, & Walker, 2006). However, defamation law is also limited in its ability regulate hate speech that is not directed at specific individuals or aimed toward causing reputational harm. Thus, hate speech does not, strictly speaking, have a legal definition.

Scholars, by contrast, have offered a variety of definitions for hate speech. For example, one popular definition in the communication literature says hate speech is "a bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics" (Cohen-Almagor, 2013, p. 43). While this definition emphasizes hostility based on innate characteristics, other definitions take it farther to emphasize speech as violence. For example, one definition describes hate speech as "a mechanism of violent subornation" that encompasses fear, harassment, intimidation, and discrimination (Lederer & Delgado, 1995). Other similar definitions highlight the potential for hate speech to incite and promote brutality and violence in a wider circle of extremists (Lederer & Delgado, 1995; Levine, 2002; Meddaugh & Kay, 2009), or the psychological harm that hate speech can cause to individuals in targeted groups (Calvert, 1997; Marwick & Lewis, 2015; Soral, Bilewicz & Winiewski, 2018). These approaches share some essential characteristics, including emphasis on expression that is discriminatory, intimidating, disapproving, antagonistic, and/or prejudicial toward specific, usually disadvantaged, groups of people and/or topics that are relevant to those groups such as race/ethnicity, gender identity, sexual orientation, religion, and disability (Cohen-Almagor, 2013; Lillian, 2007; Nemes, 2002). Additionally, these definitions conceptualize hate speech as having one of two goals: (a) conveying a message to other likeminded extremists or (b) the intimidation of targeted groups.

2.2. Exposure to hate speech in various communication settings

Expression norms and network reach are two characteristics or affordances of communication settings are important for understanding whether hate speech gets expressed and who perceives it as such, and variation in these characteristics helps us to develop theory about the settings in which the perception of hate speech is more or less likely to occur.

Expression norms shape how people approach political conversation and, ultimately, what gets said during those conversations (Eliasoph, 1998), because particular tendencies emerge within groups or networks of individuals that largely dictate whether controversial or vitriolic speech is acceptable (Barnidge, 2017). Publicness, anonymity, and group dynamics are three important factors that shape these norms. Generally speaking, people feel more comfortable expressing controversial or hostile views when communication is less public (Cowan & Hodge, 1996), or when it is public-but-anonymous (Erjavec & Kovačič, 2012). Research shows that anonymity has a "freeing" effect on expressing these views, particularly in online spaces (Davis, 1998; Chawki, 2009). Meanwhile, specific groups can establish their own particular social norms, which may encourage or discourage certain kinds of speech (Eliasoph, 1998). Hate groups are the most prominent example of groups that establish norms that are accepting or encouraging of hate speech (Duffy, 2003; Douglas, McGarty, Bliuc, & Lala, 2005).

The reach of communication is also an important factor to consider, primarily because it affects the likelihood that people will be exposed to social and/or political difference (Barnidge, 2017; Brundidge, 2010), which makes it more likely that social identity processes will result in the perception of hate speech (Leets, 2001). Larger communication networks tend to contain more weak ties, who are more likely to express social and political difference (Brundidge, 2010). Exposure to political difference is relatively more likely to prompt in-group identification and out-group differentiation (Huddy, 2001). Moreover, there are three reasons why these social identity processes may result in the perception of hate speech (Leets, 2001). First, people who belong to targeted groups may accept the negative self-image and stigmatization explicitly or implicitly communicated by hate speech. Second, individuals in targeted groups may reject these negative images and stigmatizations, applying the label "hate speech" as a way of dismissing the communication. Third, targeted groups may label communication as hate speech as a strategy for enhancing their group image as a form of social mobility.

Predictions about the frequency with which people will perceive hate speech in various communication settings can be derived from the logic outlined above. For face-to-face communication, social norms typically discourage hate speech in face-to-face conversation, where people generally seek common ground with discussion partners in an effort to reduce the potential for argument or conflict (Conover, Searing & Crewe, 2002; Eliasoph, 1998). Therefore, unless it is clear that discussion partners will share the expresser's views, the structures that govern face-to-face conversation typically limit hate speech. Meanwhile, people tend to report only a few discussion partners in face-to-face settings, with the most common type of discussant being spouses (Morey, Eveland, &

Hutchens, 2012). Therefore, face-to-face conversation tends not to have a broad reach. Based on this logic, we would expect relatively low levels of hate speech in face-to-face settings for the typical survey respondent.

By contrast, users of both social media and anonymous online message boards are more likely to perceive hate speech. Social media generally promote expression (Barnidge, Huber, Gil de Zúñiga, & Liu, 2018; Halpern & Gibbs, 2013), and they also tend to expand people's communication networks and expose them to higher levels of political difference (Barnidge, 2017; Brundidge, 2010). Meanwhile, message boards provide users with a high degree of anonymity (Davis, 1998), and they also tend to be ideologically homogeneous (Wojcieszak & Mutz, 2009), which means that people may feel more comfortable expression hateful views. However, message boards also have a broad reach (Wellman & Gulia, 1999), which means that individuals could be exposed to political difference. Therefore, to the extent that message boards expose people to political and social difference, they might promote the perception of hate speech. Based on this logic, we predict that respondents will report higher levels of hate speech in both social media and anonymous online message boards.

H1: Social media users will (a) perceive more exposure to hate speech than non-users and (b) perceive more exposure to hate speech in social media settings than in face-to-face settings.
H2: Anonymous online forum users will (a) perceive more exposure to hate speech than non-users and (b) perceive more exposure to hate speech in anonymous online forums than in face-to-face settings.

Compared to social media and anonymous online message boards, communication via mobile messaging apps is more similar to face-to-face communication in terms of both expression norms and reach. Mobile messaging tends to occur within a relatively small circle of contacts with whom the expressers typically has an offline relationship (Kim, Kim, Kim, & Wang, 2017)—that is, people they know "IRL." While the insertion of technological space between discussants may, to a small extent, reduce the need to seek common ground, the overriding tendency is to seek agreement and social harmony. Therefore, social norms do not necessarily encourage more hate speech in mobile messaging apps than they do in face-to-face settings.

RQ1: Will mobile messaging app users (a) perceive more exposure to hate speech than non-users, and (b) perceive more exposure to hate speech in mobile messaging apps than in face-to-face settings?

2.3. Avoidance of Political Talk

Exposure to extreme views in political discussions can potentially lead people to avoid political talk in the future (Eliasoph, 1998; Wells et al., 2017). For example, Wells et al., (2017) found that (a) online media users employ technological features to avoid exposure to disagreement, and (b) political discussants in face-to-face communication settings use civility and social norms to avoid unwanted political talk. Many online and mobile media give people the ability to unfriend, unfollow, hide, or block specific individuals (John & Gal, 2018; Yang, Barnidge, & Rojas, 2017), which could reduce the likelihood that people encounter unwanted speech in the future. Given that (perceived) exposure to hate speech could harm an individual's sense of self-esteem and self-worth (Calvert, 1997), these features of online/social media are an effective tool for digital-network disconnection when individuals encounter unwanted posts, including hate speech specifically aimed at disadvantaged groups. Thus, online and mobile media provide mechanisms for the "post-hoc user filtration" (Yang et al., 2017) of unwanted content. While face-to-face settings do not provide people with these same tools for preventing unwanted speech, research shows that people can reduce future exposure to such speech simply by cutting off the conversation or changing the topic (Conover et al., 2002; Eliasoph, 1998; Wells et al., 2017). Thus, people use a variety of context-specific strategies to avoid political talk with others who espouse unwanted views.

H3: Perceived exposure to hate speech will be positively related to avoidance of political talk.

3. Method

3.1. Sample and data

This study relies on a two-wave, online panel survey of adult internet users who are residents of the United States. The first wave was collected between September 19–29, 2018, six weeks before the 2018 U.S. Midterm Elections, and the second wave was collected during the month after the elections, between November 7 and December 5, 2018. The survey was administered by a private survey firm, Survey Sampling International (SSI), which used quotas based on age, gender, race, and census region. The first survey wave has a sample size of N = 1493 and a cooperation rate of 70% (AAPOR, 2016; CR3). The second survey wave has a sample size of N = 576 and a 39% retention rate. The first-wave sample is broadly reflective of the population of interest (see Table 1 for demographics and descriptive statistics for all variables).

3.2. Measures

3.2.1. Political talk avoidance.

The avoidance variables are based on recent research on the cessation of political talk (Wells et al., 2017). For face-to-face and online settings, respondents were asked whether, in the last 12 months, they have stopped talking politics with someone (1 = Yes,

Table 1
Descriptive statistics for variables in the study.

Variable Name	No. of Items	Reliability Statistics	Descriptive Statistics
Political Talk Avoidance Wave 2	10	Cronbach's alpha = 0.91	M = 1.02, SD = 2.10
Political Talk Avoidance Wave 1	10	Cronbach's alpha = 0.92	M = 1.32, SD = 2.40
Perceived Exposure to Hate Speech	16	Cronbach's alpha = 0.99	M = 2.37, SD = 1.66
Face-to-Face	4	Cronbach's alpha = 0.96	M = 2.48, SD = 1.78
Social Media	4	Cronbach's alpha = 0.97	M = 2.77, SD = 2.04
Anonymous Online Forums	4	Cronbach's alpha = 0.97	M = 2.30, SD = 1.85
Mobile Messaging Apps	4	Cronbach's alpha = 0.97	M = 2.23, SD = 1.88
Political Talk Network Size			
Face-to-Face	1		M = 13.16, SD = 24.82
Social Media	1		M = 25.43, SD = 117.64
Anonymous Online Forums	1		M = 6.72, SD = 53.29
Mobile Messaging Apps	1		M = 6.02, SD = 19.69
Political Talk Frequency			
Face-to-Face	4	Cronbach's alpha = 0.79	M = 3.68, SD = 1.53
Social Media	4	Cronbach's alpha = 0.92	M = 2.66, SD = 1.83
Anonymous Online Forums	4	Cronbach's alpha = 0.93	M = 2.19, SD = 1.68
Mobile Messaging Apps	4	Cronbach's alpha = 0.91	M = 2.61, SD = 1.79
Political Talk Diversity			
Face-to-Face	4	Cronbach's alpha = 0.86	M = 3.37, SD = 1.66
Social Media	4	Cronbach's alpha = 0.95	M = 2.62, SD = 1.87
Anonymous Online Forums	4	Cronbach's alpha = 0.95	M = 2.12, SD = 1.67
Mobile Messaging Apps	4	Cronbach's alpha = 0.94	M = 2.43, SD = 1.76
Traditional News Use	8	Cronbach's alpha = 0.83	M = 3.02, SD = 1.28
Online News Use	4	Cronbach's alpha = 0.88	M = 2.76, SD = 1.55
Social Media News Use	4	Cronbach's alpha = 0.83	M = 3.01, SD = 1.64
Mobile Messaging App News Use	1		M = 2.79, SD = 2.08
Anonymous Online Forum News Use	1		M = 2.01, SD = 1.70
Conservative Ideology	3	Cronbach's alpha = 0.95	M = 6.33, SD = 2.70
Ideological Extremity	3	Cronbach's alpha = 0.95	M = 2.09, SD = 1.75
Conservative Party Identity	3		M = -0.15, SD = 2.20
Strength of Party Identity	3		M = 1.92, SD = 1.08
Political Knowledge	6		M = 4.74, SD = 1.52
Political Efficacy	3	Cronbach's alpha = 0.70	M = 3.89, SD = 1.19
Political Interest	3	Cronbach's alpha = 0.89	M = 4.35, SD = 1.72
Age	1		M = 48.39, SD = 16.18
Gender	1		51% women
Race	1		77.2% white
Education	1		M = 4.38, SD = 1.71
Income	1		M = 4.84, SD = 2.14
Religious Affiliation	1		75% affiliated

0 = No) because of hate speech. For social media and mobile messaging apps, respondents were asked this question, and they were also asked whether they had (1) unfriended, (2) hidden, and (3) blocked and/or reported someone because of hate speech. These items were summed.

3.2.2. Perceived exposure to hate speech.

Our measures of perceived hate speech exposure are based on prior research on hate speech that takes a perceptual approach (Costello et al., 2016; Cowan & Hodge, 1996; Leets, 2001). For each of the four settings (face-to-face, social media, mobile messaging apps, and anonymous online forum), respondents were asked "In the last 12 months," how often (1 = Never, 7 = Very often) they "encounter or come across hate speech" about (1) people of a specific race or ethnicity, (2) people of a specific gender identity, (3) people of a specific sexual orientation, and (4) people of a specific religion. The four items were averaged for each of the communication settings, and an overall item was also created.

3.2.3. Political talk

Based on prior literature (Eveland & Hively, 2009), we included three dimensions of political talk as control variables: political talk network size, political talk frequency, and political talk diversity. We include a separate variable for each of the four communication settings included in the analysis: face-to-face, social media, mobile messaging apps, and anonymous online forums. Respondents were first prompted: "From time to time, people talk with others about government, elections, politics, or the news." They were then asked to enumerate how many people with whom they have "talked about these subjects" in the past 12 months in each of the four communication settings. The items were capped at 200 to reduce skew. Next, respondents were asked how often (1 = Never, 7 = Very often) they talk about "government, elections, or politics" with (1) family members, (2) friends, (3) other coworkers or classmates, and (4) other acquaintances in each of the four communication settings. The four items for each setting were averaged to create four separate variables. Finally, respondents were asked how often they talk about "government, elections, or politics" with (1)

people on the left, (2) people on the right, (3) people who have very different political views, and (4) people who have similar political views. For talk frequency and diversity, the four items for each setting were averaged to create four separate variables per dimension.

3.2.4. News use

We also include five news use variables, which are based on prior literature (Gil de Zúñiga, Jung, & Valenzuela, 2012). Traditional news use was measured with eight survey items. Respondents were asked how often (1 = Never, 7 = Several times a day), they use national newspapers, local or regional newspapers, national news magazines, national news broadcasts, local news broadcasts, cable television news, talk radio, and public radio. These eight items were averaged to create the final variable. Online news use was measured with four survey items asking respondents how often they use online-only news sites or blogs, online sites for news organizations, podcasts, and blogs. These four items were averaged. The social media news use item was based on four survey items (averaged), which asked respondents how often they use social networking websites or apps, microblogging websites or apps, photo sharing websites or apps, and video sharing websites or apps. Mobile messaging app news use was measured with a single survey item asking respondents how often they use mobile message apps for news. Anonymous online forum news use was measured with a single survey item asking respondents how often they use online message boards for news.

3.2.5. Political ideology

Based on prior literature (Garrett & Stroud 2014), political ideology was measured with three survey items asking respondents to place themselves on an 11-point, L-R scale (1 = Liberal, 5 = Neutral, 11 = Conservative) for social issues, economic issues, and general ideology. These average of these three items was taken as the final variable. The political ideology scale was folded to create an ideological extremity variable. Scores of 0 indicate moderate ideology, and 5 indicates extreme ideology.

3.2.6. Party identity

Three survey items, which were borrowed from the 2017 Annenberg National Election Study, were used to create the party identity variable. The first asked respondents, “Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what?” Those who identified as Democrat or Republican were then directed to a second question asking them how strong their identity is (Strong or Not that strong). Strong party identifiers were assigned a score of 3 (Republican) or −3 (Democrat), while weak party identifiers were assigned a score of 2 or −2. Those who identified as independents or other were directed to a different follow-up question, which asked “Even though you don’t identify with either major party, do you typically think or yourself as closer to the Democratic Party or to the Republican Party?” Those who identified as party leaners were assigned scores of 1 (Republican) or −1 (Democrat), while those who responded “Neither” were assigned a score or 0 (Non-partisan). This method resulted in a 7-point scale (−3 = Strong Democrat, 0 = Nonpartisan, 3 = Strong Republican). To create a strength of party identity variable, the above variable was folded so that 0 = Non-partisan and 3 = Strong partisan.

3.2.7. Political knowledge

Political knowledge was measured with six fact-based survey items derived from prior research (Delli Carpini & Keeter, 1996). Correct answers were tallied, with a minimum score of 0 and a maximum score of 6.

3.2.8. Political efficacy

Political efficacy was measured with three items borrowed directly from prior research (Niemi, Craig, & Mattei, 1991). Respondents were asked the extent to which they agree or disagree (1 = Strongly disagree, 7 = Strongly agree) that “People like me can influence what local government does,” “I believe that the national government cares about what people like me think,” and “City government responds to the initiatives of individuals.” These three items were averaged to create the final variable.

3.2.9. Political interest

Based on prior literature (Verba, Schlozman, & Brady, 1995), political interest was measured with three items asking respondents how interested they are in local or regional politics, national politics, and international politics (1 = Not at all, 7 = Very). These items were averaged to create the final variable (Cronbach’s alpha = 0.89, M = 4.35, SD = 1.72).

3.2.10. Demographics

Analyses also controlled for demographics, including age, gender, race, education (1 = Some high school and 7 = Post-graduate degree), income (1 = Less than $15,000 and 8 = More than $150,000), and religious affiliation.

3.3. Analysis

First, the nearest-neighbor propensity-score matching technique was combined with ANOVA-by-regression to estimate mean differences in perceived exposure to hate speech between users and non-users of various platforms (social media, mobile messaging apps, anonymous online forums). Propensity scores were constructed with a logistic regression (logit) model predicting (a) social media use, (b) mobile messaging app use, and (c) anonymous online forum use (1 = user, 0 = non-user). The nearest-neighbor method was then used to randomly match non-users to each user. Once the groups were constructed, the mean differences in exposure to hate speech were estimated through an ANOVA-by-regression (ordinary least squares [OLS]) model. These models

Table 2
ANOVA-by-regression models showing mean differences in exposure to hate speech for users versus non-users of social media, mobile messaging apps, and online forums.

Variable	Social Media B (SE)	Mobile Messaging Apps B (SE)	Online Forums B (SE)
Mean Estimates			
Intercept $_{M\ Non\text{-}Users}$	1.11 (0.10)***	1.15 (0.10)***	1.16 (0.09)***
Comparison Coefficient $\Delta_{M\ Users}$	1.21 (0.15)***	0.58 (0.10)***	1.41 (0.08)***
Control Variables			
Social Media Use	–	0.46 (0.12)***	0.51 (0.10)***
Mobile Messaging App Use	0.29 (0.14)*	–	0.26 (0.09)**
Anonymous Online Forum Use	1.54 (0.12)***	1.52 (0.09)***	–
R^2	0.41***	0.28***	0.20***
N	626	1144	1416

Notes. Cell entries are unstandardized beta coefficients (B) and standard errors (SE) from ordinary least squares regression models. Because the predictors are categorical factors, comparison coefficients can be interpreted as mean differences from the reference group (intercept). Propensity score matching was used to match users with non-users, creating three separate matched datasets for this analysis. The matching criteria included political ideology, ideological extremity, party identity, strength of party identity, political knowledge, political efficacy, political interest, age, gender, race, education, income, and religious affiliation. ***p < .001.

control for use of other mediums. Second, a repeated-measures analysis was conducted in the linear mixed effects (LME) modeling framework in order to assess mean differences in perceived exposure to hate speech among users of social media, mobile messaging apps, and anonymous online forums. These models treat individual respondents as a second-level variable with random intercepts. All control variables are mean-centered. A combined model was fit to an imputed dataset to establish a ranked order among the hate speech items. Finally, OLS regression models were used to assess the relationship between perceived exposure to hate speech and avoidance of political talk. Separate models were fit for cross-sectional and longitudinal relationships, with the latter including an autoregressive term.

4. Results

Table 2 shows results from the ANOVA-by-regression analysis. Models compare groups of users and non-users of (a) social media, (b) mobile messaging apps, and (c) anonymous online forums. For each of these models, the intercept can be interpreted as the mean of perceived exposure to hate speech for the non-users group, while the coefficient shows users' difference from that reference group mean. In the social media model, non-users have a mean of 1.11 (SE = 0.10, p < .001), while the mean for the users group is 1.21 higher (SE = 0.15, p < .001), or 2.32. Thus, the model shows that social media users perceive exposure to significantly more hate speech than non-users. This same pattern is replicated in the other two models. In the mobile messaging app model, non-users have a mean of 1.15 (SE = 0.10, p < .001), while the mean for the users group is 0.58 higher (SE = 0.10, p < .001), or 1.73. Likewise, for the anonymous online forums model, non-users have a mean of 1.16 (SE = 0.09, p < .001), while the mean for the users group is 1.41 higher (SE = 0.08, p < .001), or 2.57. These means are visualized in Fig. 1.

Table 3 shows results from the repeated-measures analysis. Once again, separate models were fit for (a) social media, (b) mobile messaging apps, and (c) anonymous online forums. Each of these models was fit using the subset of users of each platform. Because the control variables are mean-centered, the intercepts can be interpreted as the mean of the reference group (face-to-face political talk), and the comparison coefficient can be interpreted as the difference from that reference group. In the social media model, the mean for face-to-face communication is estimated as 2.48 (SE = 0.05, p < .001), while the mean for social media is 0.14 higher (SE = 0.05, p < .05), or 2.62. In the mobile messaging apps model, the mean for face-to-face communication is estimated at 2.54 (SE = 0.05, p < .001), while the mean for mobile messaging apps is 0.47 lower (SE = 0.04, p < .001), or 2.07. Finally, in the online forums model, the estimated mean for face-to-face communication is 2.35 (SE = 0.09, p < .001), while the mean for online forums is not significantly different (B = −0.02, SE = 0.07, *n.s.*). These means are visualized in Fig. 2.

While the above models provide good comparisons within relevant subgroup populations, they are not able to establish a ranked order among the means for the hate speech items. A single model is needed to establish such a rank order, and, because the data contain missing values resulting from non-use of specific platforms, it is necessary to perform a multiple imputation in order to do so. Thus, the combined model presents something of a counterfactual: What would mean levels of exposure to hate speech be on each platform *if everyone used all of the platforms*?

With that logic in mind, predictive mean matching was used to impute missing values in Wave 1. First, cases with complete data were used to predict values of variables with missing data, producing a set of coefficients. Next, a random draw was taken from the predictive posterior distribution to produce a new set of coefficients, which were then used to compute predicted values for all cases with at least one missing value. Finally, an observed value close to the predicted value of each missing case was located and assigned as a substitute. This process was repeated 50 times, and the average imputation was taken as the assigned value.

A combined repeated-measures LME model was then fit to the complete data, treating the hate speech items as a single factor

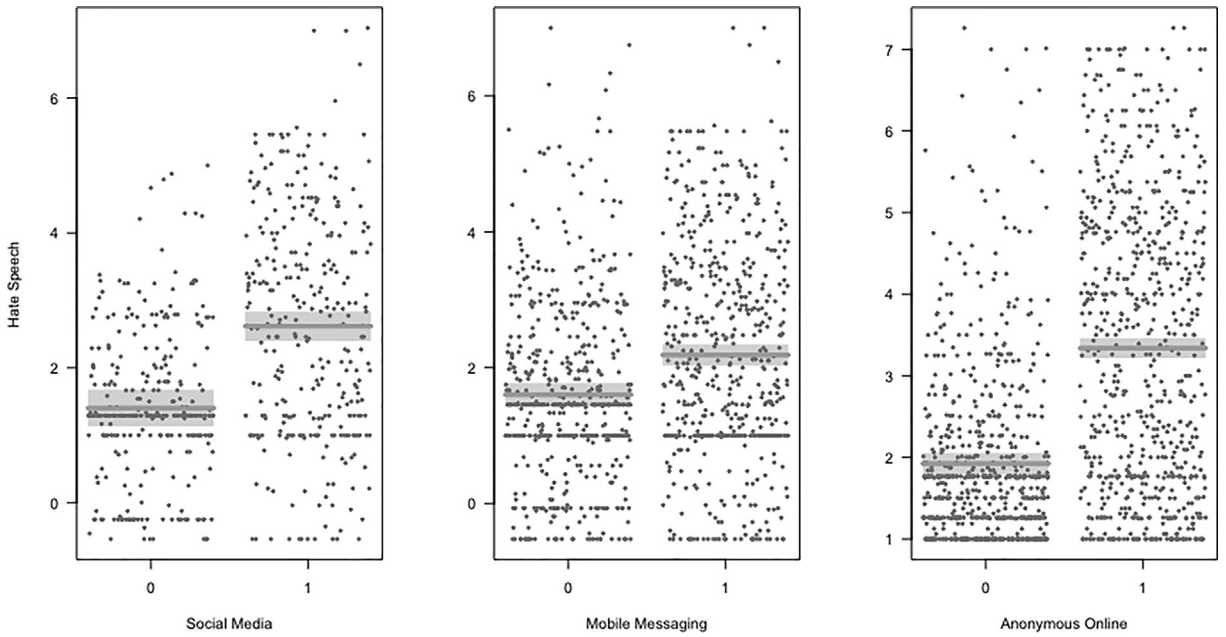

Fig. 1. Estimated mean differences in exposure to hate speech for users (1) and non-users (0) of social media, mobile messaging apps, and anonymous online forums. Means estimated from models reported in Table 2.

where face-to-face is the reference condition. Results are reported in Table 4. The model shows that social media has the highest mean, as indicated by the difference between the comparison coefficient and the intercept, at 2.63 (B = 0.14, SE = 0.05, $p < .01$). Face-to-face is second, with an estimated mean (i.e., model intercept) of 2.49 (SE = 0.05, $p < .001$). Anonymous online is third with 2.37 (B = -0.12, SE = 0.05, $p < .05$). Finally, mobile messaging apps are last with 2.00 (B = −0.49, SE = 0.05, $p < .001$). These means are visualized in Fig. 3. Mean estimates from all analyses summarized in Table 5.

All in all, these results show strong support for H1. Social media users perceive more exposure to hate speech than non-users, and they perceive it more in social media settings than in face-to-face settings. Results are mixed for H2. Anonymous online forum users perceive more exposure to hate speech than non-users; however, this may not be due to their communication in online forums, as no difference between this setting and face-to-face communication was observed. Finally, the results pertaining to RQ1 are also mixed. Group comparisons show that mobile messaging app users perceive more exposure to hate speech than non-users. However, users also report more hate speech in face-to-face settings than in mobile messaging apps.

The last set of tests H3, which predicts a positive relationship between perceived exposure to hate speech and avoidance of political talk. Table 6 shows both cross-sectional and longitudinal tests of this hypothesis. In the cross-sectional model, the slope coefficient is 0.53 (SE = 0.05, $p < .001$), indicating that for a one-unit increase in perceived exposure to hate speech, we can expect a 0.53 increase in avoidance (on a 10-point scale). In the longitudinal model, the slope coefficient is also positive, but also somewhat weaker at 0.15 (SE = 0.07, $p < .05$), which can be expected with the inclusion of the autoregressive term (B = 0.35, SE = 0.04, $p < .001$). These results provide relatively strong support for H3.

5. Discussion

The study finds that social media users perceive more exposure to hate speech than non-users, and also that they perceive more exposure to hate speech in social media environments than they do in face-to-face settings. Results are more mixed for mobile messaging apps and anonymous online forums. Comparisons of users versus non-users show, in both cases, that users perceive more hate speech than non-users. However, comparisons of settings within the subgroups of users do not follow our expectations: Results show no statistically significant difference between anonymous online forums and face-to-face communication, and they also show a lower mean for mobile messaging apps than for face-to-face communication. Finally, results show a consistent relationship between perceived exposure to hate speech and the avoidance of political talk, regardless of the communication setting.

These results point toward several concrete conclusions. First, social media tend to promote (perceived) exposure to hate speech, and, second, (perceived) exposure to hate speech is associated with the avoidance of political talk. The first conclusion echoes a growing public chorus of concern about hate speech on social media (Guiora & Park, 2017; Schmidt & Wiegand, 2017), which has been criticized for lax or non-existent self-regulation (Gopalan, 2018; Guiora & Park, 2017). Thus, the current study provides some empirical evidence validating these complaints. The second conclusion fits with several prior studies that examine the cessation of discussion (Eliasoph, 1998; Wells et al., 2017), and suggests that hate speech could be harmful for the political sphere and civil society because it could disengage people from deliberative and discursive processes that are known to promote pro-democratic and

Table 3
Repeated measures analysis showing differences in exposure to hate speech between platforms and face-to-face political talk among users of those platforms.

Variable	Social Media B (SE)	Mobile Messaging Apps B (SE)	Online Forums B (SE)
$Intercept_{M\ Face\text{-}to\text{-}Face}$	2.48 (0.05)***	2.54 (0.05)***	2.35 (0.09)***
Comparison Coefficient Δ_M	0.14 (0.05)**	−0.47 (0.04)***	−0.02 (0.07)
Face-to-Face Political Talk Network Size	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Face-to-Face Political Talk Frequency	0.01 (0.05)	−0.06 (0.04)	−0.06 (0.07)
Face-to-Face Political Talk Diversity	0.11 (0.04)**	0.08 (0.04)*	0.11 (0.06)
Social Media Political Talk Network Size	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Social Media Political Talk Frequency	0.15 (0.06)*	0.09 (0.05)	0.14 (0.07)
Social Media Political Talk Diversity	0.21 (0.05)***	0.03 (0.05)	0.00 (0.07)
Mobile Messaging App Political Talk Network Size	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Mobile Messaging App Political Talk Frequency	0.20 (0.06)***	0.22 (0.05)***	0.21 (0.08)
Mobile Messaging App Political Talk Diversity	−0.03 (0.06)	0.13 (0.05)*	0.03 (0.08)
Online Political Talk Network Size	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Online Political Talk Frequency	−0.05 (0.05)	0.02 (0.05)	−0.02 (0.07)
Online Political Talk Diversity	0.08 (0.06)	0.12 (0.05)*	0.27 (0.07)***
Traditional News Use	0.07 (0.04)	0.09 (0.04)*	0.04 (0.06)
Online News Use	0.05 (0.04)	−0.02 (0.04)	0.10 (0.05)
Social Media News Use	0.05 (0.04)	0.03 (0.04)	−0.01 (0.06)
Mobile Messaging App News Use	0.03 (0.03)	0.09 (0.03)**	0.07 (0.04)
Online Forum News Use	0.00 (0.00)	0.06 (0.03)*	−0.01 (0.04)
Party Identity (+Republican)	−0.04 (0.02)	−0.04 (0.02)	−0.02 (0.03)
Strength of Party Identity	−0.04 (0.04)	−0.08 (0.04)	−0.11 (0.06)
Political Ideology (+Conservative)	−0.02 (0.02)	−0.01 (0.01)	−0.03 (0.02)
Ideological Extremity	0.04 (0.03)	0.05 (0.02)*	0.09 (0.04)*
Political Interest	0.09 (0.03)**	0.06 (0.03)	0.09 (0.05)
Political Knowledge	−0.04 (0.03)	−0.06 (0.03)	−0.03 (0.04)
Political Efficacy	−0.04 (0.04)	−0.04 (0.04)	−0.07 (0.05)
Annual Household Income	0.00 (0.02)	0.02 (0.02)	−0.01 (0.03)
Education	−0.04 (0.03)	−0.04 (0.03)	−0.03 (0.04)
Gender Identity (1 = Woman)	−0.09 (0.09)	−0.13 (0.09)	−0.16 (0.13)
Age	−0.02 (0.00)***	−0.01 (0.00)***	−0.02 (0.01)**
Race (1 = Non-White)	0.03 (0.10)	0.00 (0.09)	0.00 (0.13)
Religious Affiliation (1 = Affiliated)	−0.16 (0.10)	−0.14 (0.09)	−0.21 (0.13)
$SD_{Intercept}$	0.80	0.70	0.78
$SD_{Residual}$	1.07	0.80	1.01
Log Likelihood	−2767.50	−2572.00	−1476.40
N	1646	1646	890
Groups	823	823	445

Notes. Cell entries are coefficients (B) and standard errors (SE) from linear mixed effects (LME) models. Data were stacked for repeated measures analysis. *p < .05, **p < .01, ***p < .001.

pro-social outcomes (Conover et al., 2002).

Third, and contrary to our hypothesis, this study found relatively lower levels of perceived hate speech in online message boards. Prior research has identified these boards as hotbeds of hate speech (Marwick & Lewis, 2015), and we expected to find relatively higher levels, but this expectation was not borne out in the findings. That said, we cannot conclude that hate speech doesn't occur on these sites, just that message board users perceive less of it. In all likelihood, hate speech occurs more frequently in some specific venues (e.g., r/The_Donald) than in others. Even where it does happen, if discussion occurs only among the likeminded, then people won't recognize hate speech for what it is. That is to say, if hate speech becomes normalized and participants agree with it, then they won't perceive it to be hate speech at all, and thus won't report it on survey questions such as the ones employed in this study. While this problem presents something of a dilemma for researchers in that it may prevent triangulation between third-party observation and self-reported perception, it also arises from the very theoretical processes outlined in this study. The perception that hate speech has occurred results not only from expression norms, but also the network reach of speech. That is, if hate speech never reaches those who may take exception to it, it will not be perceived as hate speech.

Our results also indicate relatively lower levels of perceived hate speech on mobile messaging apps. We had no concrete expectations for mobile messaging apps, because it was not clear from the prior literature that hate speech was prevalent in the context of the United States. But clearly, it is quite prevalent in other national contexts, for example in India, where the government has introduced plans to de-encrypt What's App data for the purposes of detecting and punishing hate speech offenders (Wagner, 2019). What's App is much more popular in India than in the United States, and it was one of the primary venues through which misinformation was shared during their most recent election. Policies such as these may be introduced in the United States should hate speech become more common on mobile messaging apps, including not only What's App but also Snapchat.

These conclusions can be extended to important public and scholarly conversations about the limits of free speech in the United

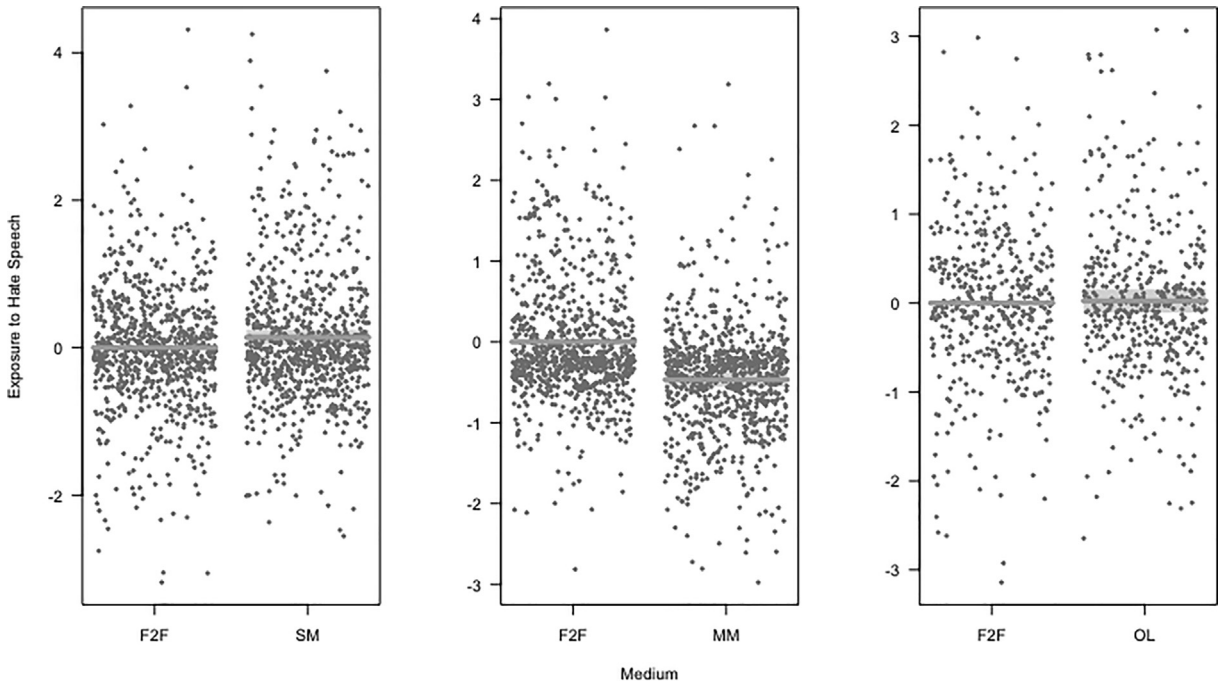

Fig. 2. Estimated mean differences in exposure to hate speech between platforms and face-to-face political talk (F2F) among users of those platforms, which include social media (SM), mobile messaging apps (MM) and anonymous online forums (OL). Means estimated from models reported in Table 3.

States. With the exception of cases of defamation, hate speech is considered protected speech under the First Amendment (Walker, 1994). The U.S. Supreme Court has ruled that several categories of speech are not protected by the First Amendment, including speech that incites violence that is imminent (for example, Schenk v. the United States [1919] established the "clear and present danger" rule, while Brandenberg v. Ohio [1969] clarifies that the threat of violence must be imminent). But hate speech is not one of those categories, and there are no laws restricting hate speech in the United States (Walker, 1994).

In the United States, public policy discussions have primarily focused on social media, most prominently including Facebook, Twitter, and YouTube. These policy recommendations largely fall into one of two camps—government regulation of social media platforms or platform regulation of themselves. Ideas that belong to the first camp largely borrow from the recent actions of foreign countries. For example, Germany's NetzDG law requires a public complaint procedure for reporting perceived hate speech (Gollatz, Riedl, & Pohlmann, 2018). Meanwhile, Australia holds internet companies and executives legally liable for hate speech (Cave, 2019).

While these types of government-centered policies also could be introduced in the United States, it seems far more likely, given the long-standing American media tradition of self-regulation (Hutchins et al., 1947), that America's largest social media corporations will regulate themselves in order to avoid government regulation. Indeed, they have already begun doing so. For example, YouTube has recently cracked down on hate speech (Brownlee, 2019), and Facebook has banned white nationalist ideas and organizations (Stack, 2019). Still, platforms could do more to combat the perception that hate speech is prevalent on their sites. For example, the Center for American Progress recommends that social media sites (a) strengthen their terms of service so that users are aware of their hate speech policies, (b) increase the transparency of algorithmic curation and content moderation, (c) develop processes for handling complaints about hate speech, as well as giving users the right to appeal hate speech complaints, and (d) continued training and evaluation of employees to more effectively deal with hate speech on their platforms (Fernandez, 2018). Each of these steps would not only limit the hate speech that occurs on social media, they would also combat the perception that it occurs, which is perhaps more important in terms of maintaining a broad user base and avoiding government regulation.

The foremost limitation of this study is that it relies on self-reported exposure to hate speech, rather than on third-party observations. Two individuals could be exposed to the same speech and come to different conclusions about whether it was hateful, and therefore there is some empirical slippage between self-reported and observational data. Future research could investigate this slippage by comparing self-reported survey measures with web-tracking data. That said, there are three arguments in favor of a perceptions-based approach. First, hate speech has no broadly accepted legal definition. Therefore, establishing criteria for making third-party observations is difficult. The self-reported approach avoids this difficulty by simply asking respondents whether they think they have been exposed to hate speech. Second, while it is true that two respondents may have different thresholds for perceiving hate speech, our study accounts for this by making within-subjects comparisons across communication settings. Thus, the study still provides valuable insight into the relative frequency of hate speech in different settings, assuming that individual over-/underestimation is consistent across settings. Third, studying perceived hate speech is valuable for understanding political behavior, because perceptions are important drivers of behavior.

Table 4
Repeated measures analysis showing differences in exposure to hate speech between platforms and face-to-face political talk.

Variable	Combined Model B (SE)
Intercept $_{M\ Face\text{-}to\text{-}Face}$	2.49 (0.05)***
Social Media Comparison Coefficient Δ_M	0.14 (0.05)**
Mobile Messaging App Comparison Coefficient Δ_M	−0.49 (0.05)***
Anonymous Online Forum Comparison Coefficient Δ_M	−0.12 (0.05)*
Face-to-Face Political Talk Network Size	0.00 (0.00)
Face-to-Face Political Talk Frequency	−0.02 (0.04)
Face-to-Face Political Talk Diversity	0.06 (0.04)
Social Media Political Talk Network Size	0.00 (0.00)
Social Media Political Talk Frequency	0.04 (0.05)
Social Media Political Talk Diversity	0.15 (0.05)**
Mobile Messaging App Political Talk Network Size	0.00 (0.00)
Mobile Messaging App Political Talk Frequency	0.21 (0.05)***
Mobile Messaging App Political Talk Diversity	0.05 (0.05)
Online Political Talk Network Size	0.00 (0.00)*
Online Political Talk Frequency	0.03 (0.05)
Online Political Talk Diversity	0.16 (0.05)**
Traditional News Use	0.06 (0.04)
Online News Use	0.01 (0.04)
Social Media News Use	0.07 (0.04)
Mobile Messaging App News Use	0.02 (0.03)
Online Forum News Use	0.03 (0.03)
Party Identity (+Republican)	−0.03 (0.02)
Strength of Party Identity	−0.05 (0.04)
Political Ideology (+Conservative)	−0.01 (0.02)
Ideological Extremity	−0.01 (0.02)
Political Interest	0.08 (0.03)**
Political Knowledge	−0.03 (0.03)
Political Efficacy	−0.07 (0.03)
Annual Household Income	0.01 (0.02)
Education	−0.03 (0.03)
Gender Identity (1 = Woman)	−0.07 (0.09)
Age	−0.02 (0.00)***
Race (1 = Non-White)	0.04 (0.09)
Religious Affiliation (1 = Affiliated)	−0.14 (0.09)
$SD_{Intercept}$	0.81
$SD_{Residual}$	0.94
Log Likelihood	−5187.70
N	3292
Groups	823

Notes. Cell entries are coefficients (B) and standard errors (SE) from linear mixed effects (LME) models. Data were stacked for repeated measures analysis. *p < .05, **p < .01, ***p < .001.

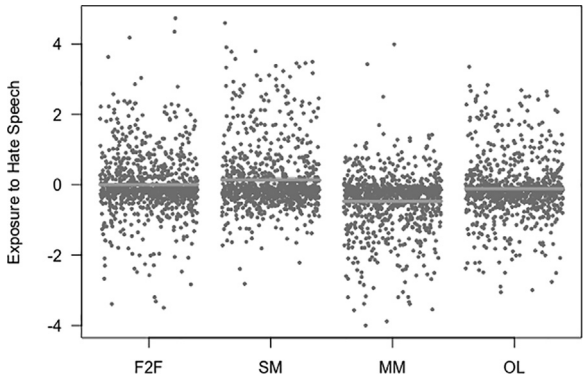

Fig. 3. Estimated mean differences in exposure to hate speech between platforms. Means estimated from the combined reported in Table 4.

Table 5
Estimated means and associated test statistics from hate speech items by type of statistical test.

	Raw Scores		Separate LME Models		Combined LME Models	
Exposure to Hate Speech …	M	t	M	t	M	t
Social Media	2.77	3.18**	2.62	2.74**	2.63	2.93**
Face-to-Face	2.48	–	2.48	–	2.49	–
Anonymous Online Forums	2.30	−4.67***	2.49	0.36	2.37	−1.98*
Mobile Messaging Apps	2.23	−11.27***	2.07	−10.59***	2.00	−2.41***

Notes. Test statistics estimated as differences from the face-to-face item. For the raw scores, test statistics from paired-samples t-tests are reported. In the LME models, test statistics from regression estimates are reported. Also in these models, means for the face-to-face items are model intercepts. M: mean, t: t-Statistic, p: p-value; LME: linear mixed effects. *p < .05, **p < .01, ***p < .001.

Table 6
The cross-sectional and longitudinal relationship between exposure to hate speech avoidance of political talk.

	Avoidance of Political Talk $_{T1}$	Avoidance of Political Talk $_{T2}$
Variable	B (SE)	B (SE)
Intercept	0.14 (0.43)	−0.54 (0.60)
Exposure to Hate Speech $_{T1}$	0.53 (0.05)***	0.15 (0.07)*
Avoidance of Political Talk $_{T1}$	–	0.35 (0.04)***
Political Talk Network Size $_{T1}$	0.01 (0.00)***	−0.01 (0.00)*
Political Talk Frequency $_{T1}$	−0.06 (0.08)	0.01 (0.11)
Political Talk Diversity $_{T1}$	0.20 (0.08)**	0.26 (0.11)*
News Use $_{T1}$	0.22 (0.07)**	0.09 (0.09)
Republican Party Identity $_{T1}$	0.00 (0.03)	−0.06 (0.05)
Strength of Party Identity $_{T1}$	−0.01 (0.06)	0.08 (0.08)
Conservative Political Ideology $_{T1}$	−0.05 (0.03)	0.02 (0.04)
Ideological Extremity $_{T1}$	0.05 (0.04)	0.02 (0.05)
Political Interest $_{T1}$	−0.06 (0.04)	−0.04 (0.06)
Political Knowledge $_{T1}$	0.06 (0.04)	−0.03 (0.06)
Political Efficacy $_{T1}$	−0.04 (0.05)	0.10 (0.07)
Annual Household Income	−0.01 (0.03)	−0.03 (0.04)
Education	−0.04 (0.04)	−0.05 (0.05)
Gender Identity (1 = Woman)	0.20 (0.13)	−0.01 (0.19)
Age	−0.01 (0.00)	0.00 (0.01)
Race (1 = Non-White)	−0.11 (0.14)	0.17 (0.19)
Religious Affiliation (1 = Affiliated)	0.00 (0.13)	−0.21 (0.18)
R^2	0.31***	0.32***
N	1481	574

Notes. Cell entries are coefficients (B) and standard errors (SE) from ordinary least squares (OLS) regression models. *p < .05, **p < .01, ***p < .001.

The study is also limited in other important ways. First, while the study relies on a longitudinal design for testing the relationship between perceived exposure to hate speech and the avoidance of political talk, readers should take caution when using these results to make causal conclusions, as the study has not eliminated all potential alternative explanations. Second, the opt-in online panel is not, strictly speaking, a true probability sample. However, research based on these samples has become increasingly common in the social sciences, and the current sample does reflect the population of interest along key criteria, including age, race, gender, and census region. A final set of limitations is related to the analysis and results. Some findings did not conform to our expectations, and the reasons for this are unknown. Specifically, we expected to observe higher levels of hate speech on anonymous online forums, such as Reddit, than in face-to-face communication, but this was not the case. Future research should investigate the occurrence of hate speech specifically in these venues, perhaps using a subset of Reddit users. Finally, while the matching procedure goes a long way toward eliminating selection bias, it cannot completely correct for it. Therefore, these results should be interpreted with caution until a true experimental design can be developed.

Despite these limitations, this study has provided relatively strong evidence that survey respondents tend to perceive a relatively high level of exposure to hate speech on social media. Furthermore, perceived exposure to hate speech is associated with the avoidance of political talk, which could have detrimental consequences for democratic societies.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

References

American Association for Public Opinion Research [AAPOR] Standard definitions: Final dispositions of case codes and outcome rates for surveys Retrieved from http://aapor.org 2016.

Barnidge, M., 2017. Exposure to political disagreement in social media versus face-to-face and anonymous online settings. Political Commun. 34 (2), 302–321. https://doi.org/10.1080/10584609.2016.1235639.

Barnidge, M., Huber, B., Gil de Zúñiga, H., Liu, J.H., 2018. Social media as a sphere for "risky" political expression: A twenty-country multilevel comparative analysis. Int. J. Press/Politics 23 (2), 161–182. https://doi.org/10.1177/1940161218773838.

Brownlee, C. (2019, June 5). YouTube's new crackdown on hate speech comes at a curious time. Slate. Retrieved from https://slate.com/technology/2019/06/youtube-hate-speech-crackdown-steven-crowder-carlos-maza.html.

Brundidge, J., 2010. Encountering "difference" in the contemporary public sphere: the contribution of the Internet to the heterogeneity of political discussion networks. J. Commun. 60 (4), 680–700. https://doi.org/10.1111/j.1460-2466.2010.01509.x.

Calvert, C., 1997. Hate speech and its harms: A communication theory perspective. J. Commun. 47 (1), 4–19. https://doi.org/10.1111/j.1460-2466.1997.tb02690.x.

Carroll, J., Karpf, D., 2018, September 22. How can social media firms tackle hate speech? University of Pennsylvania. Retrieved from http://knowledge.wharton.upenn.edu/-article/can-social-media-firms-tackle-hate-speech/.

Cave, D., 2019, April 3. Australia passes law to punish social media companies for violent posts. The New York Times. Retrieved from https://www.nytimes.com/-2019/04/03/world/australia/social-media-law.html.

Chawki, M., 2009. Anonymity in cyberspace: Finding the balance between privacy and security. Int. J. Technol. Transfer Commercialisation 9 (3), 183–199. https://doi.org/10.1504/IJTTC.2010.030209.

Cohen-Almagor, R., 2013. Freedom of expression v. social responsibility: Holocaust denial in Canada. J. Mass Media Ethics 28 (1), 42–56. https://doi.org/10.1080/08900523.2012.746119.

Conover, P.J., Searing, D.D., Crewe, I.M., 2002. The deliberative potential of political discussion. Br. J. Political Sci. 32 (1), 21–62. https://doi.org/10.1017/S0007123402000029.

Costello, M., Hawdon, J., Ratliff, T., Grantham, T., 2016. Who views online extremism? Individual attributes leading to exposure. Comput. Hum. Behav. 63, 311–320. https://doi.org/10.1016/j.chb.2016.05.033.

Cowan, G., Hodge, C., 1996. Judgments of hate speech: the effects of target group, publicness, and behavioral responses of the target. J. Appl. Soc. Psychol. 26 (4), 355–374. https://doi.org/10.1111/j.1559-1816.1996.tb01854.x.

Davis, R., 1998. The Web of Politics: The Internet's Impact on the American Political System. Oxford University Press, New York.

Delli Carpini, M.X., Keeter, S., 1996. What Americans Know About Politics and Why It Matters. Yale University Press, New Haven.

Douglas, M.K., McGarty, C., Bliuc, A., Lala, G., 2005. Understanding cyber hate: Social competition and social creativity in online white supremacist groups. Social Sci. Comput. Rev. 23 (1), 68–76. https://doi.org/10.1177/0894439304271538.

Downes, L., 2018, August 30. The summer of hate speech. The Washington Post. Retrieved from https://www.washingtonpost.com/technology/2018/08/30/summer-hate-speech/?utm_term=.edd3a3dc8cad.

Duffy, E.M., 2003. Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. J. Commun. Inquiry 27 (3), 291–312. https://doi.org/10.1177/0196859903252850.

Eliasoph, N., 1998. Avoiding Politics: How Americans Produce Apathy in Everyday Life. Cambridge University Press, Cambridge.

Erjavec, K., Kovačič, M.P., 2012. "You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. Mass Commun. Society 15 (6), 899–920. https://doi.org/10.1080/15205436.2011.619679.

Eveland Jr., W.P., Hively, M.H., 2009. Political discussion frequency, network size, and heterogeneity of discussion as predictors of political knowledge and participation. J. Commun. 59, 205–224. https://doi.org/10.1111/j.1460-2466.2009.01412.x.

Fernandez, H., 2018, October 25. Curbing hate online: What companies should do now. Center for American Progress. https://www.americanprogress.org/issues/immigration/-reports/2018/10/25/459668/curbing-hate-online-companies-now/.

Garrett, R.K., Stroud, N.J., 2014. Partisan paths to exposure diversity: differences in pro-and counter attitudinal news consumption. J. Commun. 64, 680–701. https://doi.org/10.1111/jcom.12105.

Gil de Zúñiga, H., Jung, N., Valenzuela, S., 2012. Social media use for news and individuals' social capital, civic engagement and political participation. J. Computer-Mediated Commun. 17 (3), 319–336. https://doi.org/10.1111/j.1083-6101.2012.01574.x.

Gollatz, K., Riedl, M.J., Pohlmann, J., 2018. Removal of online hate speech in numbers. Media Policy Project Blog. The London School of Economics and Political Science. Retrieved from https://blogs.lse.ac.uk/mediapolicyproject/2018/08/16/removals-of-online-hate-speech-in-numbers/.

Gopalan, S., 2018, November 16. Hate speech, fake news, privacy violations—time to rein in social media. The Hill. Retrieved from https://thehill.com/opinion/technology/417049-hate-speech-fake-news-privacy-violations-time-to-rein-in-social-media.

Guiora, A., Park, E.A., 2017. Hate speech on social media. Phiosophia 45, 957–971. https://doi.org/10.1007/s11406-017-9858-4.

Halpern, D., Gibbs, J., 2013. Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. Comput. Hum. Behav. 29 (3), 1159–1168. https://doi.org/10.1016/j.chb.2012.10.008.

Huddy, L., 2001. From social to political identity: a critical examination of social identity theory. Political Psychol. 22 (1), 127–156. https://doi.org/10.1111/0162-895X.00230.

Hutchins, R.M., Chafee Jr., Z., Clark, J.M., Dickinson, J., Hocking, W.F., Lasswell, H.D., et al., 1947. A Free and Responsible Press: A General Report on Mass Communication: Newspapers, Radio, Motion Pictures, Magazines, and Books. University of Chicago Press, Chicago.

John, N.A., Gal, N., 2018. "He's got his own sea": Political Facebook unfriending in the personal public sphere. Retrieved from. Int. J. Commun. 12, 2971–2988. https://ijoc.org/index.php/ijoc/article/view/8673/2410.

Kim, Y., Kim, B., Kim, Y., Wang, Y., 2017. Mobile communication research in communication journals from 1999 to 2014. New Media & Society 19 (10), 1668–1691. https://doi.org/10.1177/1461444817718162.

Lederer, L., Delgado, R., 1995. Introduction. In: Lederer, L., Delgado, R. (Eds.), The Price We Pay: The Case Against Racist Speech, Hate Propaganda, and Pornography. Hill and Wang, New York, pp. 3–13.

Leets, L., 2001. Explaining perceptions of racist speech. Commun. Res. 28 (5), 676–706. https://doi.org/10.1177/009365001028005005.

Levine, B., 2002. Cyber hate: a legal and historical analysis of extremists' use of computer networks in America. Am. Behav. Scientist 45 (6), 958–988. https://doi.org/10.1177/0002764202045006004.

Lillian, D., 2007. A thorn by any other name: sexist discourse as hate speech. Discourse Society 18 (6), 719–740. https://doi.org/10.1177/0957926507082193.

Lima, C., 2018, October 29. Social media's hate problem. Politico. Retrieved from https://www.politico.com/newsletters/morning-tech/2018/10/29/social-medias-hate-problem-392838.

Marwick, A., Lewis, R., 2015. Media manipulation and disinformation online. Data & Society. Retrieved from https://datasociety.net/pubs/oh/-DataAndSociety_MediaManipulationAndDisinformationOnline.pdf.

Meddaugh, P.M., Kay, J., 2009. Hate speech or "reasonable racism?" The other in Stormfront. Jo. Mass Media Ethics 24 (4), 251–268. https://doi.org/10.1080/08900520903320936.

Morey, A.C., Eveland Jr, W.P., Hutchens, M.J., 2012. The "who" matters: types of interpersonal relationships and avoidance of political disagreement. Political Commun. 29 (1), 86–103. https://doi.org/10.1080/10584609.2011.641070.

Nemes, I., 2002. Regulating hate speech in cyberspace: Issues of desirability and efficacy. Information Commun. Technol. Law 11 (3), 193–220. https://doi.org/10.1080/1360083022000031902.

Niemi, R.G., Craig, S.C., Mattei, F., 1991. Measuring internal political efficacy in the 1988 National Election Study. Am. Political Sci. Rev. 85 (4), 1407–1413. https://doi.org/10.1016/-j.chb.2013.06.005.

Peters, J., 2018, November 2. How the law protects hate speech on social media. Columbia Journalism Review. Retrieved from https://www.cjr.org/analysis/gab-hate-speech.php.
Schmidt, A., Wiegand, M., 2017. A survey on hate speech detection using natural language processing. In: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media (pp. 1-10). Retrieved from http://www.aclweb.org/-anthology/W17-1101.
Stack, L., 2019, March 27. Facebook announces new policy to ban white nationalist content. The New York Times. Retrieved from https://www.nytimes.com/2019/03/27/-business/facebook-white-nationalist-supremacist.html.
Soral, W., Bilewicz, M., Winiewski, M., 2018. Exposure to hate speech increases prejudice through desensitization. Aggressive Behavior 44 (2), 136–146. https://doi.org/10.1002/ab.21737.
Verba, S., Schlozman, K.L., Brady, H.H., 1995. Voice and Equality: Civic Voluntarism in American Politics. Harvard University Press, Cambridge.
Wagner, K., 2019, February 19. What's App is at risk in India. So are free speech and encryption. Vox. Retrieved from https://www.vox.com/2019/2/19/18224084/india-intermediary-guidelines-laws-free-speech-encryption-whatsapp.
Walker, S., 1994. Hate Speech: The History of an American Controversy. University of Nebraska Press, Lincoln, NE.
Weaver, R.L., Kenyon, T.A., Partlett, F.A., Walker, P.C., 2006. The Right to Speak III: Defamation, Reputation and Free Speech. Carolina Academic Press, Melbourne Retrieved from http://hdl.handle.net/11343/25928.
Wells, C., Cramer, K.J., Wagner, M.W., Alvarez, G., Friedland, L.A., Shah, D.V., Franklin, C., 2017. When we stop talking politics: the maintenance and closing of conversation in contentious times. J. Commun. 67 (1), 131–157. https://doi.org/10.1111/jcom.12280.
Wellman, B., Gulia, M., 1999. Net surfers don't ride alone: Virtual communities as communities. In: Smith, M.A., Kollock, P. (Eds.), Communities and Cyberspace. Routledge, New York, NY, pp. 167–194.
Wojcieszak, M.E., Mutz, D.C., 2009. Online groups and political discourse: do online discussion spaces facilitate exposure to political disagreement? J. Commun. 59 (1), 40–56. https://doi.org/10.1111/j.1460-2466.2008.01403.x.
Yang, J., Barnidge, M., Rojas, H., 2017. The politics of "unfriending": user filtration in response to political disagreement on social media. Comput. Hum. Behav. 70, 22–29. https://doi.org/10.1016/j.chb.2016.12.079.

Matthew Barnidge (Ph.D., University of Wisconsin-Madison) is an assistant professor in the Department of Journalism & Creative Media at The University of Alabama. His research specializes in emerging media and contentious political communication with an international perspective.

Bumsoo Kim (Ph.D., The University of Alabama) is a postdoctoral researcher in the Department of Communication and Journalism at The Hebrew University of Jerusalem. His research interests include digital media, political communication, and local communication ecology.

Lindsey A. Sherrill (Ph.D., The University of Alabama) is an assistant professor in the Department of Management & Marketing at the University of North Alabama. Her research interests include organizational communication, media ecology and political communication.

Žiga Luknar (M.A., University of Vienna) is a communications specialist at the European Law Institute in Vienna, Austria. He specializes in the legal and ethical aspects of hate speech.

Jiehua Zhang (M.S., Beijing Jiaotong University) is a doctoral student in the College of Communication & Information Sciences at The University of Alabama. Her research interests include political communication, digital journalism, and emerging media.