Research Center
Computer & Communications Industry Association

research.ccianet.org

2023

Hate Speech & Digital Ads: The Impact of Harmful Content on Brands

Melissa Pittaoulis

Computer & Communications Industry Association

Contents

Introduction

At present, leading U.S. social media services maintain rigorous digital trust and safety policies to protect users from harmful content.[1] Many U.S. social media services have also invested in providing advertisers and business users with detailed suitability controls to determine where and whether advertisements, digital storefronts, and other business user content appear in relation to different categories of user-generated content.[2] However, a number of "must-carry" bills have been proposed in various jurisdictions that, if enacted, could limit social media services' ability to remove or deprioritize harmful user-generated content. Two such bills recently became law in Texas and Florida, but are not yet in effect, due to pending consideration by the U.S. Supreme Court. Until this paper, there has been little public-facing research exploring the implications of hypothetical legal requirements that would require social media services to display content that would otherwise violate their current hate speech policies.

This paper examines the impact that simulated user-generated hate speech may have on consumers' perceptions of digital services' and their advertisers' brand likability and favorability. This analysis relies upon the results of two independently administered online survey experiments conducted in February and March 2023.

The studies described in this paper provide an important first step at examining the effect that unmoderated harmful content shared on social media could have on users' opinions of social media services and the brands that advertise on them. This research studied the impacts of hate speech, in particular, and found that hate speech on social media was associated with a decline in consumer sentiment towards the platform, with substantial shares of respondents reporting that such posts make them like the platform less. This finding was consistent across the three social media services tested.

In a hypothetical scenario where hate speech was not moderated on social media services, research also found negative implications for brands that advertise on the services when hate speech was viewed. Proximity to content that included hate speech resulted in some respondents reporting that the content made them like the advertiser less. It also resulted in a slight decrease in favorable opinions of the advertiser brand, as well as a larger change in net favorability, with some of the movement shifting from favorable opinions to neutral (i.e., neither favorable nor unfavorable) opinions. Respondents who viewed content with hate speech also reported a lower likelihood of purchasing the advertised brand that directly preceded the content, compared to those respondents who viewed social media content with a positive or neutral tone right after the ad.

The results suggest that consumer sentiment toward a social media service would decline if it did not remove user-generated hate speech, and that consumer sentiment would also decline for brands that advertise on the same platform adjacent to said content. These findings indicate that social media services have a rational incentive to moderate harmful content such as hate speech and are consistent with digital services' assertions that not all engagement adds value and that, in fact, some engagement is of negative value.

[1] For example, the Digital Trust & Safety Partnership, which counts leading U.S. social media services as members, describes best practices here: https://dtspartnership.org/best-practices/. Moreover, in Meta's most recent Community Standards Enforcement Report, the prevalence of hate speech was extremely low – with 1-2 views of content per 10,000 views in Q1 2023. https://transparency.fb.com/data/community-standards-enforcement/.

[2] For example, see Google content suitability controls at https://support.google.com/google-ads/answer/12764663?hl=en; Meta brand safety and suitability controls at https://www.facebook.com/business/help/1926878614264962?id=1769156093197771.

Methodology

A. Survey Administration

Survey 1 was conducted online with respondents recruited from the BizKnowledge Panel, a survey service offered by Veridata Insights, a leading market research firm. Gender and age quotas were implemented to ensure that the sample matched the U.S. adult population. A total of 1,185 respondents qualified and completed the survey. Data were collected between February 16 and March 2, 2023. To ensure that the data were of high quality, we implemented quality control procedures such as digital fingerprinting, a ReCAPTCHA test, attention check questions, timing tests (e.g., completing a survey too quickly), and straightliner exclusions (i.e., excluding respondents who select same response option in a set of questions). Survey 1 results were not weighted, as gender and age quotas were applied ex ante to ensure a sample representative of the U.S. population.

Survey 2 was conducted online by Morning Consult. A total of 2,235 respondents qualified and completed the survey. The data were collected between March 10 and March 14, 2023. In conducting the survey, Morning Consult included various quality assurance measures, similar to those used in Survey 1. These measures included procedures to prevent bots from completing the survey, timing tests, attention-check response options, and checks for straightlining grids. In reporting the results for Survey 2, Morning Consult weighted the data to match the demographics of the U.S. Census Bureau's 2022 Consumer Population Survey Annual Social and Economic Supplement. We did not observe any meaningful differences between the unweighted and weighted results. The Survey 2 results highlighted in this report are the weighted results.

B. Study Design

The surveys focused on simulated content made to appear as user-generated posts shared on three social media services: Facebook, Instagram, and Twitter. These services were chosen because they each allow their users to share images and text, thereby allowing us to create stimuli that would be amenable to testing in a survey environment.

The surveys included social media posts intended to represent three different types of posts that may appear on social media – those with a positive tone or affect, those with a neutral tone, and those with a negative tone. It should be noted that these posts were mocks – and not live examples found on any of the aforementioned services – but rather were developed as part of the survey to demonstrate content that could potentially be shared online. The posts with a negative tone were designed to simulate a hate-speech-like post.[3]

Because respondents to the surveys were volunteers, we did not want to expose them to posts that included ethnic slurs or racial epithets. Instead, we used posts that suggest affinity with a hate group by referencing the name of a hate group (the Ku Klux Klan) or hashtags used by hate groups (#14/88).

[3] The simulated negative posts did not actually appear on the tested social media services, and would likely violate digital trust and safety practices such as community standards and be removed by the social media services at present. However, the simulated negative posts could plausibly be covered by "must-carry" laws pending consideration by the U.S. Supreme Court that could prevent digital services from removing or deprioritizing such posts in the future if those laws came into effect.

Survey 1 tested three positive posts, three neutral posts, and three negative posts. Positive posts had a strong positive emotional affect associated with a celebration. By contrast, neutral posts had an intentionally neutral emotional affect, such as mentioning low-salience dates without a day off or mentioning low-engagement social events without any signal of emotional investment or personal interest from the post creator. The social media service on which the posts were simulated to appear varied, such that for each social media service included in the survey, we tested one positive post, one neutral post, and one negative post. Survey 1 also included advertisements from three different brands – an oral care brand, a home improvement retailer, and a car manufacturer. Each respondent saw one ad from each brand.

Respondents to Survey 1 were randomly assigned to one of six groups. Each group of respondents was assigned three different posts to view – one post from each of the three social media services. Of the three posts shown to an individual respondent, one had a positive tone, one had a neutral tone, and one had a negative tone. Each of the three posts also contained an advertisement for one of the three brands included in the survey. Respondents were shown the post and adjacent advertisement in a plausible facsimile of the interface or feed of a particular social media service to ensure that the presentation of the stimuli was ecologically valid. Figure 1 below shows the combinations of advertiser, social media service, and tone of post shown to each group of respondents.

Figure 1: Summary of Stimuli Used in Survey 1

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Positive						
Advertiser	Oral Care Brand	Car Manufacturer	Home Improvement Retailer	Home Improvement Retailer	Car Manufacturer	Oral Care Brand
Social Media Platform	Facebook	Instagram	Twitter	Instagram	Twitter	Facebook
Image Description	"Ready for Grandmom's 100th Birthday Party!!"	"Thank you to the person who paid for my coffee in the drive-thru this morning!!"	"National Champions!"	"Ready for Grandmom's 100th Birthday Party!!"	"Thank you to the person who paid for my coffee in the drive-thru this morning!!"	"National Champions!"
Neutral						
Advertiser	Home Improvement Retailer	Oral Care Brand	Car Manufacturer	Car Manufacturer	Oral Care Brand	Home Improvement Retailer
Social Media Platform	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram
Image Description	"Happy Arbor Day"	"Ready for the Springfield Book Club Meeting"	"Happy National Leadership Day"	"Happy National Leadership Day"	"Happy Arbor Day"	"Ready for the Springfield Book Club Meeting"
Negative						
Advertiser	Car Manufacturer	Home Improvement Retailer	Oral Care Brand	Oral Care Brand	Home Improvement Retailer	Car Manufacturer
Social Media Platform	Instagram	Twitter	Facebook	Facebook	Instagram	Twitter
Image Description	"The U.S. should ban all immigrants!!!! #loveyourrace #14/88"	"Welcome Ku Klux Klan! #loveyour race #14/88"	"Ready for the Springfield KKK Meeting"	"Welcome Ku Klux Klan! #loveyour race #14/88"	"Ready for the Springfield KKK Meeting"	"The U.S. should ban all immigrants!!!! #loveyourrace #14/88"

Survey 1 did not include all possible combinations of social media service, tone of post, and advertiser, and therefore we cannot be certain that any differences in the results observed are attributable to the tone of the post. Nevertheless, the results of Survey 1 provided important information about whether negative, hate-speech posts may have an effect on consumers' brand perceptions.

Survey 2 was conducted to further investigate the relationship between tone of post and brand perceptions. In Survey 2, each respondent was assigned to see one of nine different social media posts. The survey used a 3x3 design, in which only the tone of the post and the social media service varied. The advertiser was held constant in Survey 2: all respondents saw a social media post that included an ad for the oral care brand above it.

Although the questions asked in the surveys were largely the same, the design of the surveys differed with respect to the manner in which the questions were asked. Survey 1 used a split design in which half of respondents were asked questions about the impact of social media posts on brand likeability while the other half were asked about the impact of the posts on brand favorability. In Survey 2, all respondents were asked both the likability and favorability questions. Both surveys also included questions that assessed respondents' baseline opinions of social media services and advertisers. In other words, respondents were asked about the overall opinions of brands before being shown any social media posts. However, in Survey 1, these questions were only asked of half of the respondents, specifically, those who were asked the brand likeability questions.

Results

A. Social Media Usage

All respondents to Survey 1 reported using at least one social media service in the past 12 months. Facebook and YouTube[4] were the most commonly used services, with 84 percent of respondents reporting Facebook usage and 84 percent reporting having visited YouTube. Majorities of respondents also used Instagram (58%), while substantial minorities used Twitter (44%), Pinterest (41%), LinkedIn (33%), Snapchat (31%) and Reddit (28%).

Unlike in Survey 1, Survey 2 did not require that respondents report social media usage in the past 12 months in order to qualify for the survey. Nevertheless, 97 percent of respondents to Survey 2 reported that they had used a social media app or visited a social media site in the past 12 months. Among Survey 2 respondents, 84 percent used Facebook, 53 percent used Instagram, and 35 percent used Twitter.

B. Baseline Brand Favorability – Social Media Services

Before showing respondents any social media posts, we asked them questions intended to measure their baseline perceptions towards the social media services and advertisers tested in the survey. Respondents were instructed: *"We are going to show you the names of some major companies and brands. For each name, please indicate your overall opinion of that brand or company."* For each name, they were asked to rate their opinion of the brand as either "Very favorable," "Mostly favorable," "Neither favorable nor unfavorable,"

[4] YouTube was not included in the surveys because the video format would have required a survey design distinct from the scrolling feed of text & image posts on the tested services.

"Mostly unfavorable," or "Very unfavorable." Respondents were also given an option to indicate that they had never heard of the brand and an option to select "Don't know."

Among respondents to Survey 1, about half reported that their overall opinion of Facebook (53%) and Instagram (48%) was mostly or very favorable, while Twitter was viewed favorably by about 28 percent of respondents. In general Twitter had higher unfavorable ratings, with 36 percent of respondents reporting that their overall opinion of Twitter was either mostly or very unfavorable. In contrast, Facebook and Instagram were viewed unfavorably by 25 percent and 15 percent, respectively.

In Survey 2, the same question was asked, but the scale was changed, with the "Mostly Favorable" and "Mostly Unfavorable" options changed to "Somewhat Favorable" and "Somewhat Unfavorable." Among Survey 2 respondents, 59 percent reported that their overall opinion of Facebook was very or somewhat favorable, while 49 percent and 33 percent reported favorable opinions of Instagram and Twitter, respectively. Like Survey 1, respondents in Survey 2 reported higher unfavorable opinions of Twitter (31%) than Facebook (23%) and Instagram (18%).

C. Baseline Brand Favorability – Advertisers

The three advertisers used in Survey 1 varied in their favorability, although none of the three were generally viewed unfavorably. The oral care advertiser was viewed favorably by 76 percent of respondents and unfavorably by just 4 percent of respondents. Twenty percent of respondents said their overall opinion of the oral care brand was neither favorable nor unfavorable. In contrast to the oral care advertiser, respondents were more likely to report neutral opinions of the home improvement retailer and car manufacturer and less likely to report favorable views. The home improvement retailer was viewed favorably by 74 percent of respondents and unfavorably by 8 percent of respondents, with 17 percent reporting that their overall opinion of this brand was neither favorable nor unfavorable. The corresponding percentages for the car manufacturer were 54 percent favorable, 5 percent unfavorable, and 32 percent neither favorable nor unfavorable.

Among Survey 2 respondents, 77 percent reported that their overall opinion of the oral care brand was either very favorable or somewhat favorable. Only 4 percent of respondents reported their overall opinion of this brand as either very or somewhat unfavorable.

D. Impact of Social Media Post on Brand Likeability

In Table 1 below, we show the results from Survey 1 of the impact of the social media post on brand likeability. As shown in this table, after being exposed to a mock social media post featuring hate speech (labeled "Negative" post in the tables), respondents were significantly more likely to report that the post made them like the social media service less, compared to when they were exposed to posts with either positive or negative tones. On average, when respondents viewed a positive or neutral post, 5 percent reported that the post made them like the social media service less. In contrast, an average of 40 percent of respondents reported liking the social media service less after viewing the negative, hate speech post. The results were similar across services – 37 percent of respondents reported that the post made them like Instagram less, 42 percent reported that it made them like Twitter less, and 41 percent reported that it made them like Facebook less.

Table 1: Impact of Tone of Post on Likability of Social Media Services – Survey 1

	Positive Post			Neutral Post			Negative Post		
	Group 1	**Group 2**	**Group 3**	**Group 1**	**Group 2**	**Group 3**	**Group 1**	**Group 2**	**Group 3**
	Grandmom birthday	Paid for coffee	National Champions	Happy Arbor Day	Springfield Book Club	National Leadership Day	Ban all immigrants #loveyourrace	Welcome KKK sign	KKK Meeting
	Facebook	Instagram	Twitter	Twitter	Facebook	Instagram	Instagram	Twitter	Facebook
It makes me like them more	11.7%	20.3%	6.1%	8.2%	14.2%	9.6%	3.1%	9.1%	3.0%
It makes me like them less	7.1%	5.1%	6.6%	3.6%	4.6%	4.1%	36.7%	41.6%	40.6%
It does not affect how I feel about them	76.5%	72.1%	81.7%	79.6%	79.7%	82.7%	52.6%	46.7%	53.3%
Don't know	4.6%	2.5%	5.6%	8.7%	1.5%	3.6%	7.7%	2.5%	3.0%
	100%	100%	100%	100%	100%	100%	100%	100%	100%
Number of Respondents	196	197	197	196	197	197	196	197	197

Source: NERA Survey, Q3a

In Table 2, we show the impact of the tone of the post on likability of social media services as measured in Survey 2. Consistent with the results of Survey 1, we see that respondents are more likely to say that a negative post makes them like the social media service less. Across all social media services, 37 percent of respondents reported that the negative post made them like the social media service less, compared to 6 percent who saw a positive post and 7 percent who saw a neutral post. The results were similar for the three social media services tested. Among respondents who viewed a negative post on Facebook, 35 percent reported that the post made them like Facebook less. Among respondents who viewed a negative post on Instagram, 40 percent reported that the post made them like Instagram less. And among respondents who viewed a negative post on Twitter, 35 percent reported that the post made them like Twitter less.

Positive posts did not produce a corresponding effect on likeability in Survey 2. In other words, compared to those who saw a neutral or negative post, seeing a positive post did not result in a substantial share of respondents indicating that they liked the social media service more.

Table 2: Impact of Tone of Post on Likability of Social Media Services – Survey 2

Survey Question:
“Thinking again about this type of post – the one below the ad – which of the following best describes how this type of post makes you feel about [Social Media Platform]?”

	Positive Post			**Neutral Post**			**Negative Post**		
	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter
It makes me like [Social Media Platform] more	16%	19%	11%	15%	14%	14%	12%	9%	10%
It makes me like [Social Media Platform] less	4%	5%	10%	9%	6%	6%	35%	40%	35%
It does not affect how I feel about [Social Media Platform]	66%	61%	61%	60%	63%	57%	39%	40%	39%
Don’t know	14%	14%	19%	16%	17%	23%	15%	11%	16%
	100%	99%	101%	100%	100%	100%	101%	100%	100%
Number of Respondents (Weighted)	219	234	255	236	244	277	236	274	261

Note: Columns may not sum to 100% due to rounding.
Source: Morning Consult Survey, CCIA 6

Next, we looked at the impact of the tone of the post on consumers' perceptions toward advertisers. In Table 3, we show the results from Survey 1 and see that the impact on advertisers is smaller than the impact on social media services. After being exposed to a hate speech post, an average of 12 percent of respondents reported that the post made them like the advertiser less, whereas after they viewed a positive or neutral post, an average of 3 percent said the post made them like the advertiser less.

Table 3: Impact of Tone of Post on Likability of Advertiser – Survey 1

	Positive Post			Neutral Post			Negative Post		
	Group 1	**Group 2**	**Group 3**	**Group 1**	**Group 2**	**Group 3**	**Group 1**	**Group 2**	**Group 3**
	Grandmom birthday	Paid for coffee	National Champions	Happy Arbor Day	Springfield Book Club	National Leadership Day	Ban all immigrants #loveyour race	Welcome KKK sign	KKK Meeting
	Oral Care Brand	Car Manufacturer	Home Improvement Retailer	Home Improvement Retailer	Oral Care Brand	Car Manufacturer	Car Manufacturer	Home Improvement Retailer	Oral Care Brand
It makes me like them more	12.8%	18.8%	11.7%	23.0%	16.8%	15.7%	6.6%	10.7%	9.6%
It makes me like them less	3.1%	3.6%	4.1%	1.5%	3.6%	3.0%	7.7%	12.2%	16.2%
It does not affect how I feel about them	81.1%	76.6%	82.7%	70.9%	76.6%	77.7%	77.6%	72.6%	71.6%
Don't know	3.1%	1.0%	1.5%	4.6%	3.0%	3.6%	8.2%	4.6%	2.5%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%
Number of Respondents	196	197	197	196	197	197	196	197	197

Source: NERA Survey, Q2a

In Table 4, we show the results for Survey 2 of the impact of the tone of the post on the likability of the oral care advertiser. Among respondents who viewed a negative post, 20 percent reported that the post made them like the advertiser less. The results were similar for each social media service tested. Among those who viewed the negative post on Facebook, 19 percent said that it made them like the advertiser less. Among those who viewed it on Instagram, 22 percent said that it made them like the advertiser less. And among those who viewed the negative post on Twitter, 19 percent said that it made them like the advertiser less.

Table 4: Impact of Tone of Post on Likability of Advertiser – Survey 2

Survey Question:
"Which of the following best describes how this type of post – the one below the ad – makes you feel about [Oral Healthcare Brand]?"

	Positive Post			**Neutral Post**			**Negative Post**		
	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter
It makes me like them more	25%	20%	16%	26%	19%	22%	13%	10%	14%
It makes me like them less	1%	5%	4%	4%	4%	2%	19%	22%	19%
It does not affect how I feel about them	67%	64%	72%	61%	62%	63%	54%	55%	60%
Don't know	7%	11%	8%	10%	14%	13%	14%	12%	8%
	100%	100%	100%	101%	99%	100%	100%	99%	101%
Number of Respondents (Weighted)	219	234	255	236	244	277	236	274	261

Notes: Columns may not sum to 100% due to rounding.
Source: Morning Consult Survey, CCIA 4

E. Impact of Social Media Post on Brand Favorability

Survey 1 finds that the negative posts produce shifts in the favorability of the social media services; however, the results are not uniform. Compared to the positive and neutral posts, the negative posts produced no difference in favorable or unfavorable opinions of Facebook. In contrast, compared to posts with positive and neutral tones, respondents report lower favorability and higher unfavorable opinions of Instagram after seeing a negative, hate-speech post on Instagram. Whereas 53.5 percent of respondents who viewed a positive post on Instagram reported that they had a very or mostly favorable view of Instagram, this percentage is only 40.9 percent for respondents who were exposed to the negative post. In addition, while only 9.6 percent reported an unfavorable opinion of Instagram after seeing a positive post, 19.7 percent reported an unfavorable opinion after viewing the hate-speech post. For Twitter, the negative post produced an increase in unfavorable opinions. While 29.8 and 23.2 percent of respondents reported that they had unfavorable opinions of Twitter after seeing the positive and neutral posts, respectively, this percentage increases to 39.7 percent for respondents who viewed a negative post on Twitter.

Table 5: Impact of Tone of Post on Brand Favorability of Social Media Services – Survey 1

	Positive Post			**Neutral Post**			**Negative Post**		
	Group 4	**Group 5**	**Group 6**	**Group 4**	**Group 5**	**Group 6**	**Group 4**	**Group 5**	**Group 6**
	Grandmom birthday	Paid for coffee	National Champions	National Leadership Day	Happy Arbor Day	Springfield Book Club	Welcome KKK sign	KKK Meeting	Ban all immigrants #loveyourrace
	Instagram	Twitter	Facebook	Twitter	Facebook	Instagram	Facebook	Instagram	Twitter
Very favorable	17.7%	12.1%	14.6%	14.6%	9.6%	17.1%	14.6%	13.6%	12.1%
Mostly favorable	35.9%	18.7%	31.7%	19.7%	38.9%	32.2%	28.3%	27.3%	17.6%
Neither favorable nor unfavorable	32.8%	34.3%	24.1%	35.9%	28.3%	32.2%	26.3%	33.3%	26.1%
Mostly unfavorable	5.1%	18.2%	14.6%	15.2%	13.1%	10.6%	18.2%	9.1%	20.6%
Very unfavorable	4.5%	11.6%	12.6%	8.1%	9.1%	4.0%	9.6%	10.6%	19.1%
Never heard of	0.5%	0.0%	0.0%	1.0%	0.0%	1.5%	0.5%	0.5%	1.0%
Don't know	3.5%	5.1%	2.5%	5.6%	1.0%	2.5%	2.5%	5.6%	3.5%
	100%	100%	100%	100%	100%	100%	100%	100%	100%
Favorable	53.5%	30.8%	46.2%	34.3%	48.5%	49.2%	42.9%	40.9%	29.6%
Unfavorable	9.6%	29.8%	27.1%	23.2%	22.2%	14.6%	27.8%	19.7%	39.7%
Number of Respondents	198	198	199	198	198	199	198	198	199

Source: NERA Survey, Q6

Table 6 shows the results from Survey 2. Similar to Survey 1, compared to the positive and neutral posts, the negative post appeared to have little impact on the favorability of Facebook, but was associated with a decrease in favorability for Instagram and an increase in unfavorable opinions of Instagram and Twitter.

Table 6: Impact of Tone of Post on Brand Favorability of Social Media Services – Survey 2

Survey Question:
"Which of the following best describes your opinion of [Social Media Platform]?"

	Positive Post			Neutral Post			Negative Post		
	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter
Very favorable	23%	25%	15%	31%	25%	18%	27%	19%	15%
Somewhat favorable	27%	23%	13%	26%	26%	13%	21%	17%	13%
Neither favorable nor unfavorable	23%	29%	32%	16%	24%	26%	26%	23%	22%
Somewhat unfavorable	13%	9%	16%	14%	9%	15%	13%	16%	16%
Very unfavorable	12%	6%	11%	9%	7%	13%	9%	18%	19%
Never heard of company or brand	0%	1%	1%	0%	0%	0%	0%	0%	2%
Don't know/no opinion	3%	8%	13%	4%	9%	15%	5%	7%	13%
	101%	101%	101%	100%	100%	100%	101%	100%	100%
Favorable	50%	48%	28%	57%	51%	31%	48%	36%	28%
Unfavorable	25%	15%	27%	23%	16%	28%	22%	34%	35%
Number of Respondents (Weighted)	219	234	255	236	244	277	236	274	261

Notes: Columns may not sum to 100% due to rounding.
Source: Morning Consult Survey, CCIA 9

Because Survey 2 included a pre-post test design for overall favorability, an alternate analysis is to compare the respondents' opinions of the social media service after seeing the negative post to the opinion they reported having before they were exposed to the post. The net favorability is the percentage of respondents whose opinion of the brand was favorable minus the percentage who had an unfavorable opinion of the brand. Before being shown a negative post on Instagram, the net favorability of Instagram was 27 percent (48% favorable – 21% unfavorable). After being shown the negative post, the net favorability of Instagram decreased to 2 percent (36% favorable – 34% unfavorable), a decline of 25 percentage points. The change in the net favorability of Instagram was smaller for the positive posts declining just 3 percentage points for the positive post; there was no change in net favorability for the neutral post. Before being shown a negative post on Twitter, the net favorability of this service was -2 percent (34% favorable – 36% unfavorable). After being shown the negative post, the net favorability of Twitter decreased to -7 percent (28% favorable – 35% unfavorable).

In Table 7, we show the impact of the social media posts in Survey 1 on brand favorability of advertisers. For the home improvement store and car manufacturer, we observe no difference in brand favorability based on the tone of the user-generated post that appears below the advertiser's post. However, for the oral care brand, there is an observed difference. While 66 percent of respondents reported a favorable view of the oral care brand after seeing the positive post and 70 percent reported a favorable view after viewing a neutral post, only 57 percent reported a favorable view of the brand after viewing a negative post. Given that the oral care brand had high baseline favorability, these results could suggest that brands with the highest favorability may be especially vulnerable to negative impact on brand perception caused by proximity to user-generated hate speech posts. Further research is needed to determine whether this effect replicates with other product types or is specific to the particular oral care advertisement utilized in the research.

Table 7: Impact of Tone of Post on Brand Favorability of Advertisers – Survey 1

	Positive Post			**Neutral Post**			**Negative Post**		
	Group 4	**Group 5**	**Group 6**	**Group 4**	**Group 5**	**Group 6**	**Group 4**	**Group 5**	**Group 6**
	Grandmom birthday	Paid for coffee	National Champions	National Leadership Day	Happy Arbor Day	Springfield Book Club	Welcome KKK sign	KKK Meeting	Ban all immigrants #loveyour race
	Home Improvement Retailer	Car Manufacturer	Oral Care Brand	Car Manufacturer	Oral Care Brand	Home Improvement Retailer	Oral Care Brand	Home Improvement Retailer	Car Manufacturer
Very favorable	28.8%	18.7%	21.6%	24.7%	29.8%	27.6%	22.2%	27.8%	19.1%
Mostly favorable	42.9%	40.4%	44.2%	32.3%	40.4%	46.2%	34.8%	43.9%	42.7%
Neither favorable nor unfavorable	19.7%	32.8%	26.6%	34.8%	24.7%	17.6%	32.3%	17.2%	30.2%
Mostly unfavorable	6.1%	3.5%	3.5%	4.0%	2.5%	3.5%	5.1%	3.0%	5.0%
Very unfavorable	1.0%	1.0%	3.0%	2.0%	1.0%	2.5%	3.5%	5.6%	1.5%
Never heard of	0.5%	0.5%	0.0%	0.5%	0.5%	0.5%	0.5%	0.0%	0.5%
Don't know	1.0%	3.0%	1.0%	1.5%	1.0%	2.0%	1.5%	2.5%	1.0%
	100%	100%	100%	100%	100%	100%	100%	100%	100%
Favorable	71.7%	59.1%	65.8%	57.1%	70.2%	73.9%	57.1%	71.7%	61.8%
Unfavorable	7.1%	4.5%	6.5%	6.1%	3.5%	6.0%	8.6%	8.6%	6.5%
Number of Respondents	198	198	199	198	198	199	198	198	199

Source: NERA Survey, Q5

Survey 2 further explores the impact of the tone of the post on brand favorability and finds a small change in brand favorability as a result of the tone of the social media post that appears below the ad. Regardless of the social media service on which the ad and user-generated post appeared on, among those who viewed a positive post, 5 percent viewed the oral care brand unfavorably (i.e., either as “somewhat unfavorable” or “very favorable”). For those who viewed a neutral post, 3 percent viewed the brand unfavorably. These results are similar to the baseline results, which found that 4 percent of respondents viewed the oral care brand unfavorably. In contrast, among those who viewed a negative post, 13 percent reported an overall unfavorable opinion of the oral care brand. The results for the specific social media services are shown below in Table 8.

As with the social media services, we also looked at changes in net favorability for the oral care brand. Before being exposed to a negative social media post, 78 percent of respondents reported a favorable opinion of the oral care brand, compared to 4 percent who reported an unfavorable opinion. As a result, the net favorability of the oral care brand was 78-4=74 percent. After viewing a negative post, the net favorability of the oral care brand was 46 percent (59 percent favorable vs. 13 percent unfavorable), a decline of 24 percentage points. Some of the movement in favorability was a shift into the “neither favorable nor unfavorable” category. Such large shifts were not observed among those who viewed positive or neutral posts. The change in net favorability for positive posts was -9 percent, while the change for neutral posts was -5 percent.

Table 8: Impact of Tone of Post on Brand Favorability of Advertiser – Survey 2

Survey Question:
“Which of the following best describes your opinion of [Oral Healthcare Brand]?”

	Positive Post			**Neutral Post**			**Negative Post**		
	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter
Very favorable	40%	41%	33%	40%	38%	34%	28%	28%	33%
Somewhat favorable	32%	26%	31%	34%	35%	34%	27%	31%	29%
Neither favorable nor unfavorable	20%	22%	26%	18%	17%	23%	25%	24%	16%
Somewhat unfavorable	2%	4%	3%	3%	3%	1%	8%	5%	9%
Very unfavorable	2%	3%	1%	0%	2%	2%	5%	8%	6%
Never heard of company or brand	0%	1%	0%	0%	0%	0%	1%	0%	1%
Don’t know/ no opinion	4%	4%	6%	4%	5%	6%	6%	3%	6%
	100%	101%	100%	99%	100%	100%	100%	99%	100%
Favorable	72%	67%	64%	74%	73%	68%	55%	59%	62%
Unfavorable	4%	7%	4%	3%	5%	3%	13%	13%	15%
Number of Respondents (Weighted)	219	234	255	236	244	277	236	274	261

Notes: Columns may not sum to 100% due to rounding.
Here, “Oral Care Brand” is a pseudonym for the actual brand name tested.
Source: Morning Consult Survey, CCIA 8

F. Reaction to Post

In Survey 1, we included a question to measure respondents' reactions to the user-generated posts. This question asked how respondents would respond to seeing the user-generated post and included one non-reactive response ("keep scrolling") and four reactive actions ("react or comment," "hide post," "report post," and "ignore post but post yourself"). This question also served as a way to check whether respondents had paid attention to the content of the user-generated posts. As shown in Table 9 below, the hate speech posts generated greater reactions than the positive or neutral posts. On average, the hate-speech posts were more likely to result in a reactive action than the positive and neutral posts. In addition, they were also more likely to result in a post being reported. For example, while less than 1 percent of respondents indicated that they would report the positive or neutral posts, 38 percent of respondents indicated they would report the KKK meeting post, 33 percent said they would report the Welcome KKK post, and 17.1 percent indicated that they would report the "Ban all immigrants" post. Higher shares of respondents also indicated that they preferred not to see such posts – with an average of 11 percent indicating they would hide the hate speech post.

Table 9: Reaction to Post – Survey 1

	Positive Post			Neutral Post			Negative Post		
	Group 4	**Group 5**	**Group 6**	**Group 4**	**Group 5**	**Group 6**	**Group 4**	**Group 5**	**Group 6**
	Grandmom birthday	Paid for coffee	National Champions	National Leadership Day	Happy Arbor Day	Springfield Book Club	Welcome KKK sign	KKK Meeting	Ban all immigrants #loveyour race
Keep scrolling	43.4%	57.1%	65.8%	62.6%	71.7%	58.8%	28.3%	31.3%	40.7%
React or comment	46.0%	32.8%	18.1%	20.2%	17.2%	33.2%	15.7%	17.2%	18.1%
Hide post	1.0%	1.0%	2.0%	1.5%	1.0%	1.0%	14.1%	8.6%	9.0%
Report post	0.0%	0.5%	1.0%	0.5%	0.5%	0.5%	32.8%	37.9%	17.1%
Ignore post but post yourself	2.5%	1.5%	2.0%	2.0%	0.5%	0.5%	1.0%	1.0%	5.0%
None of these	5.6%	6.1%	8.5%	10.6%	5.6%	5.5%	5.6%	2.0%	8.0%
Don't know	1.5%	1.0%	2.5%	2.5%	3.5%	0.5%	2.5%	2.0%	2.0%
	100%	100%	100%	100%	100%	100%	100%	100%	100%
Number of Respondents	198	198	199	198	198	199	198	198	199

Source: NERA Survey, Q4

In Survey 2, the question was changed so that rather than ask about what the respondents would do regarding the mock, user-generated social media post, the question asked how they would respond to the advertisement shown above the user-generated post. The question asked, "Which of the following comes closest to what you would do in response to seeing this ad above this type of social media post?" Response options included "Keep scrolling," "React or comment on the ad," "Hide the ad," "Report the ad," and "Click the ad." Regardless of which social media service the ad appeared on, the share of respondents who reported that they would click on the ad was generally low; however, these shares were even smaller for those who viewed a negative post. For example, while 14 percent and 12 percent of respondents who viewed positive and neutral posts, respectively, indicated they would click on the ad, the share of respondents who viewed a negative post and indicated they would click on an ad was 9 percent. Respondents who saw negative posts were also more likely to indicate that they would report an ad, compared to those who viewed positive and neutral posts. Whereas almost no respondents reported that they would report an ad after seeing a positive (0%) or neutral post (1%), 20 percent of respondents who viewed a negative post indicated they would report the ad. Results for the specific services tested are shown below in Table 10.

Table 10: Reaction to Post – Survey 2

Survey Question:
"Which of the following comes closest to what you would do in response to seeing this ad above this type of social media post?"

	Positive Post			**Neutral Post**			**Negative Post**		
	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter
Keep scrolling	53%	49%	64%	56%	49%	54%	31%	39%	44%
React to or comment on the ad	13%	11%	8%	8%	8%	9%	8%	6%	8%
Hide the ad	4%	2%	1%	4%	4%	3%	9%	8%	7%
Report the ad	1%	0%	0%	1%	1%	1%	24%	21%	16%
Click the ad	15%	17%	12%	12%	15%	8%	7%	10%	10%
None of these	7%	17%	10%	8%	13%	14%	11%	11%	11%
Don't know/ no opinion	8%	4%	5%	10%	9%	11%	11%	5%	5%
	101%	100%	100%	99%	99%	100%	101%	100%	101%
Number of Respondents (Weighted)	219	234	255	236	244	277	236	274	261

Notes: Columns may not sum to 100% due to rounding.
Source: Morning Consult Survey, CCIA 7

G. Purchase Likelihood

Survey 2 included a question about likelihood of purchase that was not included in Survey 1. This question asked:

When thinking about this type of social media post – the one below the ad – does this type of post make you more or less likely to purchase a product from [oral care brand]? Or does this type of post have no effect on your likelihood to buy a product from [oral care brand].

The results of this question suggest that the tone of social media posts may have an impact on consumers' likelihood of purchasing brands whose advertisements appear in proximity to posts containing hate speech. Nearly one-quarter (23 percent) of respondents reported they are "less likely to buy" a product from the oral care brand due to the negative post shown, compared to only 6 percent and 5 percent who had seen a positive or neutral post, respectively. The results did not vary based on the social media service on which the post and ad appeared.

Table 11: Impact of Tone of Post on Brand Favorability of Advertiser – Survey 2

Survey Question:
"When thinking about this type of social media post – the one below the ad – does this type of post make you more or less likely to purchase a product from [oral care brand]? Or does this type of post have no effect on your likelihood to buy a product from [oral care brand]."

	Positive Post			**Neutral Post**			**Negative Post**		
	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter	Facebook	Instagram	Twitter
More likely to buy product	21%	19%	18%	23%	22%	20%	10%	12%	13%
No impact on my likelihood to buy product	69%	69%	70%	64%	65%	66%	55%	59%	55%
Less likely to buy product	6%	7%	4%	4%	6%	6%	25%	23%	22%
Don't know/ no opinion	5%	5%	7%	9%	7%	9%	11%	7%	9%
	101%	100%	99%	100%	100%	101%	101%	101%	99%
Number of Respondents (Weighted)	219	234	255	236	244	277	236	274	261

Notes: Columns may not sum to 100% due to rounding.
Source: Morning Consult Survey, CCIA 10

Discussion

The studies described in this paper provide an important first step at examining the potential impact of "must-carry" policies, if applied to social media services, by evaluating the effect that hate speech posted on social media may have on social media users' opinions of social media services and the brands that advertise on them. This research finds that simulated hate speech on social media is associated with a decline in consumer sentiment towards the service, with substantial shares of respondents reporting that such posts make them like the service less. This finding was consistent across the three social media services tested.

Results on brand favorability were more mixed. While simulated hate speech posts were associated with a decrease in net favorability for Instagram, there appeared to be no impact on brand favorability of Facebook. Twitter saw the share of unfavorable opinions rise slightly when hate speech posts were shown.

Social media posts that include hate speech may also have negative implications for brands that advertise on the services where the hate speech is visible. Proximity to a post that included hate speech resulted in some respondents reporting that the post made them like the advertiser less. It also resulted in a slight decrease in favorable opinions of the advertisers' brand, as well as a larger change in net favorability, with some of the movement shifting from favorable opinions to neutral (i.e., neither favorable nor unfavorable) opinions. Respondents who viewed a post with simulated hate speech before or after an advertisement also reported a lower likelihood of purchasing the advertised brand, compared to those respondents who viewed a social media post with a positive or neutral tone before or after an advertisement.

Further research is needed to determine whether hate speech has an effect on brand perceptions of social media services other than Facebook, Instagram, and Twitter. Since this experiment was conducted using synthetic still images or graphics, testing hate speech in videos or other user-generated formats is an area that future research could also pursue. In addition, studies should expand the pool of sample advertisers and the types of harmful content that legal requirements may require services to display to test whether the effects on consumers' perceptions of advertisers varies.

The opinions expressed herein do not necessarily represent the views of NERA Economic Consulting or any other NERA consultant. Please do not cite without explicit permission from the author.