Toxic Speech and Collective Engagement: Evidence from Roe v. Wade Fallout

Romilla Syed
University of Massachusetts Boston
Romilla.Syed@umb.edu

Emmanuelle Vaast
McGill University
Emmanuelle.Vaast@mcgill.ca

Abstract

This study examines how toxicity in connective action movements impacts the public's collective engagement that manifests through interactions on social media postings. We use the case of the 2022 Roe v. Wade fallout and examine the toxicity in the Instagram image and text postings made by pro-life and pro-choice groups. Our analysis focuses on four temporal events encompassing the pre- and post-Roe v. Wade eras. The results suggest that while both groups post toxic content, the toxicity is more in images than in text postings. Further, toxicity in text reduces interactions, although the patterns vary between the two groups across the temporal events and content type. This study contributes to connective action literature by providing insights into toxic speech in opposing movements. The findings might inform social media platforms to design better techniques and processes for detecting and demoting toxicity that would otherwise deter the pursuit of social movements.

Keywords: toxicity, connective action, social media, collective engagement, Roe v. Wade.

1. Introduction

Toxic speech that is "rude, disrespectful, or unreasonable language that is likely to make someone leave a discussion"[1] has become pervasive in online conversations. While in the extant literature, several terms have been used to imply offensive discourse, including hate speech, abuse, harassment, and trolling, these terms indicate that the speech is targeted towards an individual or group and violates platform service terms. In contrast, toxicity implies the use of language that may or may not be specifically targeted toward an individual or a group but is regarded as disrespectful in a broader societal context and independent of any platform or publishers' interpretation.

In this study, we are interested in analyzing toxicity in connective action movements that view the role of social media as central to organizing and mobilizing action (Bennett & Segerberg, 2012, 2013). In this grassroots-led movement, multiple actors related by their shared political interests collectively engage in furthering the movement through the use of social media features such as likes, retweets, shares, or comments. Thus the connective action is considered successful if a large number of actors collectively "come together spontaneously and informally, even if they do not all equally identify with a common cause and engage in coparticipation and coproduction of content with the use of social media" (Vaast et al., 2017, p. 1180).

However, connective action movements are often challenged by countermovements, as was witnessed, for instance, in the cases of #MeeToo and #BlackLivesMatter that led to the emergence of #MenToo and #AllLivesMatter. Through the use of social media, these opposing movements organize and seek civic participation by making competing claims and vying for public attention to effect or resist a social change (Gallagher et al., 2018; Meyer & Staggenborg, 1996). In this "tango of mobilization and demobilization" (Zald & Useem, 1987, p. 247), participants often generate and spread hateful and divisive narratives (Bharati et al., 2019; Shahin, 2023). Scholars argue that the lack of governance structures and policies around offensive content creates a "toxic technoculture," as was evidenced in the cases of #Gamergate and #TheFappening, anti-feminist and misogynistic movements on Reddit (Massanari, 2017).

An emerging body of scholarship examines the role of social media in creating, scaling, and sustaining connective action movements (Syed & Silva, 2023; Young et al., 2019). However, much of the existing studies have focused on a single movement, and far less is known about how toxicity in opposing movements manifest on social media and how that impacts collective engagement. The potential of social media to seed and amplify toxic and hateful speech has raised unprecedented challenges, especially for its disproportionate impact on the vulnerable and

---

[1] Toxicity | Jigsaw (google.com)

HICSS

marginalized members of society. A recent study by the Anti-Defamation League (2021) notes that women, LGBTQ+, religious minorities, and people of color are more likely to experience online harassment. Further, the study found that the most common response of the target to online harassment is to stop, reduce or change online behavior, including posting less often and avoiding certain sites. In the case of connective action movements, toxicity could lead to digital repression as participants perceive an increased cost for pursuing movement activity on social media (Earl et al., 2022). Thus, if online toxicity silences voices and discourages collective engagement, it not only impedes the pursuit of movement goals but also has dire consequences for freedom of expression and reduces the diversity of thought in online conversations. To that end, we aim to answer the following research question: *How does toxic speech in connective and counter-connective action movements impact collective engagement on social media?*

We use the case of Roe v. Wade fallout marked by the leak of the US Supreme Court opinion to overturn Roe V. Wade, followed by the Court's landmark ruling on June 24, 2022, that rolled back the legal right to an abortion in the US. This case is ideal for analyzing toxicity, for the fallout is expected to impact women and people of color more (Delaney, 2022), who, as noted before, are more likely to experience online harassment. We collected data from Instagram for *pro-choice* – those who support leaving the decision to abort a pregnancy up to the pregnant person and *pro-life* – those who oppose abortion altogether. The data included text and image postings, as toxic speech can be conveyed through different forms of expression[2]. Informed by a computationally intensive theory construction paradigm (Berente et al., 2019; Miranda et al., 2022), we adopted a multi- and mixed-methods approach to data analysis to identify patterns in toxicity and collective engagement across four temporal events that encompass pre- and post-Roe v. Wade eras.

The result suggests that both pro-choice and pro-life postings contained toxicity; however, the toxicity in image postings is more than in text postings. The pro-choice group had more interactions after the Supreme Court's draft opinion leaked, and the interactions declined post-fallout. The pro-choice groups had the highest interactions in the fallout and post-leak periods. We also observed a strong association between toxicity attributes and interactions. However, compared to image posting, text posting with toxicity led to reduced interactions,

although the patterns varied during the four time periods. To that end, this study contributes to connective action research that has so far been focused on a single movement and provides evidence of how toxicity in connective-counterconnective movement impacts collective engagement. We also contribute at the empirical level by examining the content of the images and contrasting it with text postings. As our results suggest, the content of the images can portray more toxicity than text postings. The findings have implications for social media platforms and social movement activists. Social media platforms need to consider designing better techniques and processes for toxicity detection and moderation and preventing toxic content promotion. Social movement activists must remain wary of toxicity that could otherwise deter the pursuit of justice and push people back to silence.

2. Background literature and theory

In this section, we first conceptualize (counter) connective action and collective engagement. Next, we discuss the existing connective action research and how toxicity in connective action impacts collective engagement.

2.1. (Counter)connective action and collective engagement

Social media plays an important role in organizing social movements by empowering and enabling citizens to facilitate and coordinate actions and bring a change in their community (Tye et al., 2018). Conceptualized as connective action, such movements represent collective engagement through the use of digital and social platforms as multiple users spontaneously and informally engage in the coproduction and cosharing of content (Bennett & Segerberg, 2012). The users assume interdependent roles by creating and amplifying content related to the cause by generating hashtags and memes, liking, sharing, or commenting (Vaast et al., 2017). Thus, connective action emerges through the use of social media features as users "commit to an action" and "recommend it to others" (Bennett & Segerberg, 2013, p. 16).

Social movement scholarship suggests that movements are often challenged by countermovements (Ayoub & Chetaille, 2020). As Inata (2021) argues, "one group's excessively large mobilization promotes counter-mobilization by another seeking to maintain the status quo" (p. 2). Both

---

[2] What is hate speech? | United Nations

movements and countermovements share the same concern; however, they make competing claims and vie for public attention to mobilize support in their favor and demobilize the supporters of the opposition (Meyer & Staggenborg, 1996). Often, these opposing movements create meta-contrasting narratives drawing boundaries between "us" v. "them" using an overly positive image of self and negative images of the opposition (McGarty et al., 1993). In recent years, in the US, several movements and countermovements have been mobilized especially on social media, on issues related to abortion, women's rights, and gun control (Gallagher et al., 2018; Shahin, 2023). Thus, it is reasonable to conclude that a successful connective action movement will likely evoke a counterconnective action. The face-off between the two is what we argue manifests toxicity on social media and impacts collective engagement, as discussed in section 2.3.

2.2. Connective action and social media

This study examines how toxicity in connective-counterconnective action impacts collective engagement. An emerging body of information systems literature has begun to examine the role of social media in organizing and mobilizing connective action movements (see Syed & Silva, 2023; Young et al., 2019). We examine the existing scholarship to understand 1) the focal phenomenon being studied and 2) the empirical context or settings being analyzed.

In terms of the focal phenomenon, prior studies have examined the use of social media platforms for developing collective sensemaking (Oh et al., 2015), the emergence of protest cycles (Tarafdar & Ray, 2021), relaxation of structural constraints and content hegemonizing (Miranda et al., 2016), raising awareness (McKenna, 2020), empowerment mechanisms (Leong et al., 2019), and driving environmental sustainability (Tim et al., 2018). Overall the use of social media features allows users to generate and amplify content (Vaast et al., 2017), develop a more organized collective action (Leong et al., 2020), and provide long-term movement sustainability (Syed & Silva, 2023). A few studies have also examined the negative dynamics of online activism. For example, through challenges to social movement organization's values (Selander & Jarvenpaa, 2016), deinstitutionalization of social norms and behaviors (Park et al., 2021), and trolling behaviors (Bharati et al., 2019).

Related to empirical settings, studies have analyzed various types of content, including tweets, posts, pictures, videos, and petitions on digital platforms such as Twitter, Facebook, YouTube, and Weibo (Tarafdar & Ray, 2021). Others have examined social media affordances to render interactions through the use of features such as hashtags, mentions, URLs, likes, shares, and comments (Selander & Jarvenpaa, 2016; Vaast et al., 2017). A few studies have also elicited data through interviews and observations (e.g., Leong et al., 2020; Leong et al., 2019; Selander & Jarvenpaa, 2016). Furthermore, the aforementioned studies have examined various research contexts, ranging from politically motivated ones, such as the Egyptian Revolution and the Stop Online Piracy Act, to environmental movements, such as the Gulf of Mexico Oil Spill, and environmental sustainability and more recently, women or feminist movements such as MeeToo, New Delhi rape case, and Womens March. Yet others examined digital action from an organizational perspective, such as Amnesty International and student organizations.

In summary, while the existing research has enhanced our understanding of the role of social media in organizing digital action, there are three gaps. *First*, much of the existing research has focused mainly on the use of social media features to generate, amplify, or sustain digital action. However, the negative dynamics associated with digital action have not received much attention despite the calls made in the scholarship (Tim et al., 2018). *Second*, much of the existing literature has focused on a single movement, and little attention has been paid to opposing movements. As noted before, contemporary digital movements are often challenged by opposing movements, especially when the object of concern is highly debatable, such as abortion, gun control, or other politically-motivated movements (Gallagher et al., 2018; Shahin, 2023). *Third*, while existing research has examined data from several digital platforms, the focus has been on textual content, including postings, keywords, and hashtags. However, little attention has been paid to the content of the images that "convey and construct powerful political meanings" (Sutton & Vacarezza, 2020, p. 733) and effectuate movement mobilization (Halfmann & Young, 2010). Furthermore, while activists disseminate both text and images, the prevalence of visual imagery to represent feminist issues, including abortion, and to emphasize dissent and confrontation between pro-choice and pro-life supporters has been well noted (Rohlinger & Klein, 2012).

Informed by the aforementioned gaps, in this study, we examine the negative dynamics in connective and counterconnective movements by focusing on toxicity in text and image postings. As noted before, toxic and hate speech on social media is pervasive. Particularly competing movements create narratives portraying the opposition as evil or immoral

to mobilize support in their favor and demobilize opposition. Such toxic speech, if it reduces the public's collective engagement, would impede the pursuit of social movement goals, as we discuss in the next section.

2.3. Toxicity and digital repression

Digital and social media platforms have proven to be important tools for advancing the goals of social movements and activists, as well as for the surveillance and suppression of movement actors (George & Leidner, 2019; Nurik, 2022). In the social movement scholarship, the countervailing forces on social media have been studied under the concept of "digital repression," which is defined as "actions directed at a target to raise the target's costs for digital social movement activity and/or the use of digital or social media to raise the costs for social movement activity, wherever that contestation takes place" (Earl et al., 2022). Substantively, digital repression is a broader concept that includes but is not limited to the use of traditional repressive techniques such as harassment, disinformation, and surveillance against digital activists or protestors. Furthermore, digital repression can not only deter the efforts of activists to pursue the movement's goals but also impact the boarder audience of sympathizers. For instance, Weidmann and Rød (2019) argue that digital intervention and surveillance negatively affect protest emergence. In other instances, digital repression could deter activists and core members of a cause but could also backfire by mobilizing new supporters (Earl & Beyer, 2014; Pan & Siegel, 2020).

Following Earl et al. (2022), in this study, we argue that toxic speech on social media could lead to digital repression if it impedes collective engagement in furthering the cause and goals of a movement. Movements are considered successful if they can mobilize a large number of people toward their cause and goals. Social media alleviates authorship constraints, allowing diverse participants to engage in discursive practices and mobilize action (Miranda et al., 2016). However, it is also prudent that the relaxed constraints coupled with a lack of governance structures and policies around toxic speech create an ideal environment for participants with different ideologies to engage in vilification and trolling of opposers (Massanari, 2017). Toxic speech that deters the efforts of activists or supporters by increasing the "cost" of engaging with the movement can have far-reaching consequences for activism and social mobilization. As noted before, toxicity implies the use of language that is regarded as disrespectful in a broader societal context and is likely to make someone leave the conversation. Toxic speech is independent of any platform or publishers' interpretation as the terms and policies vary by platform. Furthermore, toxic speech includes but is not limited to "hate speech" or "abuse" that is targeted toward an individual or a group based on their inherent characteristics, such as race, religion, or gender.

3. Research methodology

Informed by computationally intensive theory-building paradigm (Berente et al., 2019; Miranda et al., 2022), we adopt multiple and mixed-methods approach to data analysis to identify emergent patterns of toxicity and collective engagement. Figure 1 summarizes our research approach.
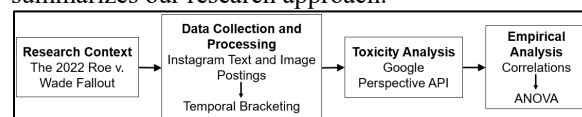

Figure 1. Research approach

3.1. Research context: The 2022 Roe v. Wade Fallout

Our context is informed by the decision of the US Supreme Court on June 24, 2022, to overturn the 1973 verdict in Roe v. Wade, which enshrined the legal right to an abortion in the US. Through a majority vote, the justices decided that abortion was not a constitutional right and returned the issue to the states to decide how to regulate abortion. The decision triggered multiple protests and debates across the US (New York Times, 2022). Social media reactions to the official ruling on certain platforms, including Twitter, Instagram, and TikTok, were reported to rise to billions of postings and views (Edwards, 2022). Furthermore, before the official ruling, the leak of the Court's draft opinions led supporters and opposers of abortion rights to take to the streets and mobilize on social media (Chang et al., 2023).

Roe v. Wade is an interesting context to analyze for its historical and political significance. Abortion in the US remains a highly polarized and politically charged issue, and Americans continue to be divided along the lines of pro-choice and pro-life (Brenan, 2021). Scholars relate Roe v. Wade and its fallout to a broader reproductive justice movement that builds on the legacy of women-led movements for bodily sovereignty and social justice (Fixmer-Oraiz, 2022). The term reproductive justice was coined in 1994 by Black feminists and claims three tenets: the right to have a child, the right not to have a child, and the right to parent one's child(ren). In doing so, it links abortion and birth control struggles to environmental justice,

gun violence, and LGBTQ+ rights, to name a few.

However, activists engaged in social movements often face vitriolic treatment, which could splinter social justice work (Ross, 2019). People on both sides of the issue engage in callouts and shaming of the opposition publicly on social media (Bharati et al., 2019; Shahin, 2023). The loss of Roe has led activists to make renewed calls to organize for rights and justice, as some argue the decision is especially consequential for the young, poor, immigrant, and women of color (Delaney, 2022). Given that many citizens and activists use social media platforms to express their views and mobilize collective action, it is prudent that factions who are against the outcome will utilize social media to organize, especially in the post-fallout period. At the same time, we expect anti-abortion groups to continue to use social media to uphold the ruling. We intend to examine the toxicity in the social media postings of two groups and how that impacts collective engagement.

3.2. Data collection, processing, and analysis

Given our research context focuses on abortion rights that weigh more on younger populations, including children and teens (Lantos et al., 2022), we choose to collect data from Instagram, as 73% of US teens use it[3]. 47.8% of all Instagram users are female, while 52.2% are male. 31% of users are between 25-34 years old, and 31% are between 18-24 years old. Furthermore, 71.90% of posts on Instagram are images. We collected Instagram posts for eight months, from December 1, 2021, to July 31, 2022. We used Meta's CrowdTangle platform to collect the data. We used two sets of search terms to represent pro-choice and pro-life factions. Pro-choice search terms included #pro-choice, #reproductiverights, #abortionright, #bansoffourbodies, and #abortionishealthcare. Pro-life search terms included #pro-life, #RoeReversal, #abortionisgenocide, #LifeWins, and #EndAbortion. Our dataset includes posts from public Instagram accounts that are verified and "influential" users with more than 50K followers, such as celebrities, politicians, non-profits, public figures, and the like. The data contained both text postings and image content. Further, we retrieved the number of interactions for each post (i.e., the sum of likes, comments, and shares), which allowed us to analyze the collective engagement for a post (Vaast et al., 2017).

As we are interested in examining toxicity and collective engagement patterns over time, we divided the data into four periods that mark the pre- and post-Roe v. Wade eras. Specifically, the temporal bracketing was informed by two significant events: 1) leaked draft of the US Supreme Court to overturn Roe v. Wade, published on May 2, 2022. 2) the official ruling of the US Supreme Court on June 24, 2022. Accordingly, we defined four periods: 1) *Pre-leak* - December 1, 2021, to May 1, 2022. 2) *Post-leak* - May 2, 2022, to June 23, 2022. 3) *Fallout* - June 24, 2022, to July 1, 2022, and 4) *Post-fallout* - July 2, 2022, to July 31, 2022.

Next, we calculated the toxicity scores for Instagram posts using Google Perspective API. Each post is scored along six toxicity attributes, as described in Table 1. The API provides a probability score for each toxicity attribute between 0 and 1. The higher the probability score, the more likely a reader would perceive the post as containing a specific toxicity attribute. Since our Instagram posts include content from both text and image postings, we computed the toxicity scores for both data types separately.

Table 1. Toxicity Attributes

Toxicity Attribute and Definition	Example excerpt
Toxicity: A rude, disrespectful, or unreasonable comment likely to make people leave a discussion	Fuck You SCOTUS! Today is a devastating day for Women's Right's, … for America!
Severe toxicity: A very hateful, aggressive, disrespectful comment or otherwise very likely to make a user leave a discussion or give up on sharing their perspective	Pro-life my fucking ass. Let's talk about all the native kids you've killed!!
Identity attack: Negative or hateful comments targeting someone based on their identity.	SOYBOYS GETTING VASECTOMIES TO IMPRESS FEMINISTS …. ENDING THEIR LINEAGE
Insult: Insulting, inflammatory, or negative comments towards a person or a group	DONT VOTE EAT TRASH DO DRUGS GET CASH EAT ASS DIE FAST
Profanity: Swear words, curse words, or other obscene or profane language	No one has authority over my body but myself and fuck you if you think you do. "Keep your fucking opinions away from MY body"
Threat: Comment describes an intention to inflict pain, injury, or violence against an individual or group	Abortion. Is. Murder. … KILLS AS MANY BLACK PEOPLE EVERY FOUR DAYS AS THE KLAN KILLED IN 150 YEARS

Finally, we conducted the correlation analysis of pro-life and pro-choice data sets to understand the relation between toxicity attributes and collective engagement. As noted before, we used total interactions (i.e., the sum of likes, comments, and shares) for a posting to indicate collective

---

[3] 2023 Instagram Statistics // Everything You Need to Know from SMPerth

engagement. Thus, we checked the correlation between scores of toxicity attributes and interactions for the posts. We conducted the analysis for both text and image postings for four time periods. We followed up with ANOVA tests to determine whether there is a statistically significant difference in the means of interactions between text and image postings. We also conducted a two-way ANOVA to test the joint effect of toxicity attributes and posting types (i.e., image or text) on interactions during the four periods. We present the results in the next section.

4. Empirical results

The descriptive statistics for pro-choice data are presented in Table 2, and pro-life in Table 3. Figures 2 and 3 present a visual representation of the toxicity attributes for the two groups across four temporal events.
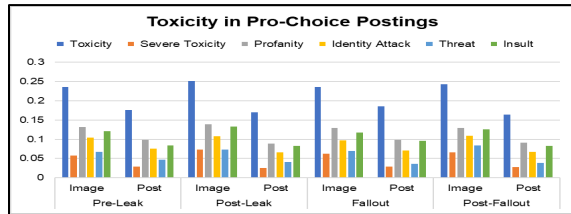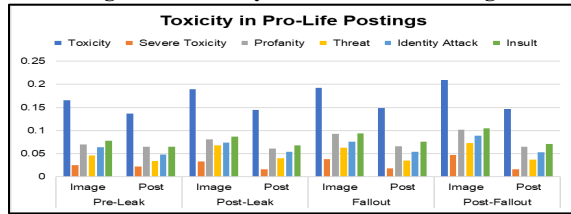

Figure 2: Toxicity in Pro-Choice Postings

Figure 3: Toxicity in Pro-Life Postings

Table 2. Descriptive statistics for pro-choice

Attributes	Mean	SD	Min	Max
Text postings (n=7317)				
Toxicity	0.17	0.17	0.00	0.99
Severe toxicity	0.03	0.08	0.00	0.92
Profanity	0.09	0.15	0.00	0.99
Identity attack	0.07	0.11	0.00	0.88
Insult	0.09	0.13	0.00	0.87
Threat	0.04	0.08	0.00	0.66
Image postings (n=3174)				
Toxicity	0.24	0.21	0.00	0.99
Severe toxicity	0.06	0.15	0.00	1.00
Profanity	0.13	0.19	0.00	1.00
Identity attack	0.10	0.15	0.00	0.98
Insult	0.12	0.17	0.00	0.96
Threat	0.07	0.12	0.00	0.96

Table 3. Descriptive statistics for pro-life

Attributes	Mean	SD	Min	Max
Text postings (n=10061)				
Toxicity	0.14	0.16	0.00	0.97
Severe toxicity	0.02	0.07	0	0.98
Profanity	0.06	0.11	0.01	0.99
Identity attack	0.05	0.09	0	0.78
Insult	0.07	0.12	0.00	0.85
Threat	0.04	0.07	0	0.96
Image postings (n=6632)				
Toxicity	0.18	0.17	0	0.99
Severe toxicity	0.03	0.09	0	0.98
Profanity	0.08	0.12	0	0.99
Identity attack	0.07	0.11	0	0.98
Insult	0.08	0.13	0	0.96
Threat	0.06	0.17	0	0.99

Table 4. Descriptive statistics for interactions

	Mean	SD	Min	Max	Count
Pro-choice					
Full data	1658.32	7315.45	0	310061	8133
Pre-SC leak	2058.30	5634.90	0	78578	1991
Post-SC leak	1686.66	7300.92	0	139570	2607
Fallout	1644.47	10039.49	0	310061	2147
Post-fallout	1052.77	3392.86	0	40827	1388
Pro-life					
Full data	1140.73	3806.51	0	90343	10953
Pre-SC leak	1031.45	2719.05	0	53440	5588
Post-SC leak	1286.77	4860.98	0	90343	2717
Fallout	1405.08	5463.59	0	84628	1309
Post-fallout	1042.03	3226.91	0	40827	1339

The pro-choice dataset contained 8133 postings, with 7317 (90%) text and 3174 (39%) image postings. The pro-life dataset contained 10953 postings with 10061 (92%) text and 6632 (61%) image postings. For pro-choice, on average, both text and image posting contained more toxicity, followed by insult, profanity, identity attack, threat, and severe toxicity. For pro-life, on average, both text and image posting contained more toxicity, followed by profanity, insult, identity attack, threat, and severe toxicity. However, for both pro-choice and pro-life data, the probability scores for the six toxicity attributes are more for image postings than text postings. Table 4 presents the summary statistics for interactions. For pro-choice, the average number of interactions decreased as the four periods unfolded, with the highest interactions in the pre-leak period and the least in the post-fallout period. In comparison, for pro-life, the highest interaction was in the fallout period, followed by post-leak, post-fallout, and pre-leak periods.

Tables 5 and 6 summarize the results of the correlation tests between the toxicity attributes and interactions. Overall, for both pro-choice and pro-life, the correlation between toxicity attributes and interactions is weak but positive. However, the correlation is more for image postings than text postings. For pro-choice, the correlation between toxicity attributes and interactions remains more for images during pre- and post-leak periods. The correlations for most toxicity attributes remain stronger for image postings in the other two periods as well. However, during the fallout period, the

correlation between profanity and interactions is more for text postings. Similarly, in the post-fallout period, the correlations of severe toxicity, identity attack, and threats with interactions are stronger for text postings. We found similar patterns for the pro-life dataset. The correlation between toxicity attributes and interactions is more for the images in the pre-leak period. The same is true for the post-leak period, except that the correlation of toxicity with interactions is stronger for text. Likewise, in the fallout period, the correlation of toxicity attributes with interactions is stronger for images except for identity attack, which is stronger for text. Finally, in the post-fallout period, the correlations of all toxicity attributes with interactions are stronger for text except for the insult attribute, which is stronger for image postings.

Table 5. Correlation test for interactions in pro-choice

	Tox.	Sev. Tox.	Prof.	Iden. Atk.	Insult	Threat
Full data						
Text	0.07	0.02	0.08	0.08	0.03	0.00
Image	0.16	0.08	0.15	0.14	0.15	0.03
Pre-leak						
Text	0.16	0.02	0.22	0.24	0.08	-0.04
Image	0.29	0.15	0.28	0.31	0.28	0.07
Post-leak						
Text	0.05	0.01	0.03	0.04	0.03	0.01
Image	0.15	0.10	0.16	0.14	0.15	0.04
Fallout						
Text	0.03	0.02	0.04	0.00	0.00	0.02
Image	0.07	0.02	0.04	0.02	0.03	0.03
Post-fallout						
Text	0.13	0.08	0.13	0.17	0.10	0.07
Image	0.19	0.06	0.17	0.07	0.19	0.00

Table 6. Correlation test for interactions in pro-life

	Tox.	Sev. Tox.	Prof.	Iden. Atk.	Insult	Threat
Full data						
Text	0.10	0.01	0.06	0.07	0.07	0.03
Image	0.11	0.06	0.08	0.06	0.10	0.08
Pre-leak						
Text	0.13	-0.00	0.07	0.09	0.10	0.01
Image	0.16	0.05	0.11	0.11	0.14	0.08
Post-leak						
Text	0.07	-0.01	0.01	0.02	0.06	-0.00
Image	0.04	0.05	0.03	0.02	0.06	0.06
Fallout						
Text	0.05	0.06	0.06	0.04	0.03	0.12
Image	0.12	0.09	0.08	0.03	0.05	0.17
Post-fallout						
Text	0.17	0.15	0.23	0.19	0.13	0.08
Image	0.16	0.08	0.16	0.07	0.18	0.02

Tables 7 and 8 present the ANOVA results for pro-choice and pro-life datasets to test for the mean interactions for image and text postings. Generally, toxicity attributes positively affect interactions for both groups; however, the results vary by content type (i.e., images or text) and temporal periods. For both datasets in the pre-leak period, compared to image postings, text postings led to reduced interactions (-624.68 for pro-choice and -188.61 for pro-life). However, the p-value is not significant during other periods. Thus, the null hypothesis that the means of

interactions for image and text posting are equal cannot be rejected. Further, for pro-choice, as shown in Table 7, the two-way ANOVA results suggest that toxicity, severe toxicity, insult, and threat in text postings have statistically significant decreasing effects on interactions in the pre-leak period. Text postings with identity attacks have a statistically significant and increasing effect on interactions in the post-fallout period. As shown in Table 8, for pro-life, text postings with toxicity, severe toxicity, profanity, insult, and threat have a statistically significant negative effect on interactions in the pre-leak period. Text postings with severe toxicity, profanity, and identity attack have a statistically significant positive effect on interactions in the pre-fallout period.

Table 7. ANOVA for pro-choice (DV: Interactions)

	Coefficients (P-Value)			
Phase (n)	Pre-leak (2786)	Post-leak (3269)	Fallout (2654)	Post-fallout (1782)
Type (Ref: Text (7317); Image=3174)	-624.68 (0.009**)	148.83 (0.586)	-35.645 (0.93)	-153.53 (0.384)
Toxicity	9322.39 (0.000***)	3926.66 (0.040*)	2655.45 (0.040*)	3203.98 (0.000***)
Toxicity*Type	-3969.72 (0.001**)	-1903.47 (0.162)	-1136.50 (0.561)	-646.99 (0.467)
Severe Toxicity	7664.42 (0.000***)	3689.32 (0.038*)	1145.53 (0.395)	1375.44 (0.003**)
SevereToxicity *Type (Text)	-5846.92 (0.012*)	-3064.06 (0.203)	601.80 (0.860)	1874.42 (0.221)
Profanity	10280.08 (0.000***)	4756.10 (0.000***)	1479.82 (0.0423*)	3202.23 (0.000***)
Profanity*Type	-989.87 (0.498)	-3166.19 (0.044*)	991.25 (0.644)	-255.88 (0.803)
Identity Attack	14203.17 (0.000***)	5286.90 (0.000***)	1367.21 (0.619)	1401.40 (0.000***)
IdentityAttack *Type	-1629.91 (0.365)	-1985.40 (0.348)	-956.04 (0.766)	3914.11 (0.001**)
Insult	11703.96 (0.000***)	4884.27 (0.000***)	125.94 (0.572)	3884.70 (0.000***)
Insult*Type	-8078.03 (0.000***)	-3321.11 (0.05)	-1335.45 (0.602)	-1354.40 (0.228)
Threat	4527.88 (0.537)	1808.76 (0.416)	2115.92 (0.319)	24.434 (0.089)
Threat*Type	-7011.56 (0.003 **)	-1072.40 (0.696)	201.74 (0.964)	2875.94 (0.069)

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05

Table 8. ANOVA for pro-life (DV: Interactions)

	Coefficients (P-Value)			
Phase (n)	Pre-leak (8869)	Post-leak (4016)	Fallout (1855)	Post-fallout (1953)
Type (Ref: Text (10061); Image=6632)	-188.61 (0.001**)	-181.82 (0.262)	-122.37 (0.607)	-40.95 (0.783)
Toxicity	3046.39 (0.000***)	1228.07 (0.000***)	317.11 (0.000***)	2729.00 (0.000***)
Toxicity*Type	-984.05 (0.007**)	944.04 (0.357)	-1913.66 (0.169)	793.02 (0.348)
Severe Toxicity	2051.04 (0.032*)	2958.10 (0.185)	4881.82 (0.002**)	2184.68 (0.000***)
SevereToxicity *Type (Text)	-2132.47 (0.003**)	-4205.00 (0.055)	-659.63 (0.828)	6731.60 (0.000***)
Profanity	3101.81 (0.000***)	1167.35 (0.268)	3320.76 (0.001**)	3335.97 (0.000***)
Profanity*Type	-1603.54 (0.003**)	-812.68 (0.588)	-607.40 (0.752)	4058.02 (0.000***)
Identity Attack	3113.50	1050.36	1135.15	1684.08

	(0.000***)	(0.136)	(0.123)	(0.000***)
IdentityAttack *Type	-419.40 (0.493)	100.37 (0.950)	937.76 (0.664)	5215.11 (0.000***)
Insult	3447.31 (0.000***)	2308.14 (0.000***)	2037.42 (0.100)	3748.34 (0.000***)
Insult*Type	-1225.59 (0.010*)	75.92 (0.953)	-1096.83 (0.532)	-318.48 (0.769)
Threat	2678.69 (0.000***)	2853.13 (0.038*)	9146.60 (0.000***)	493.96 (0.039*)
Threat*Type	-2355.31 (0.000***)	-3143.95 (0.065)	-1719.51 (0.515)	2767.84 (0.062)

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05

5. Discussion and conclusions

In this study, we examine the toxic speech generated by pro-choice and pro-life supporters and how that impacts collective engagement on social media. Informed by empirical insights and theoretical foundations, we forward the concept of *connective repression* that underscores the negative impact of repressive digital actions, such as toxicity, on connective action movements. Our analysis focused on six different forms of toxic repression – toxicity, severe toxicity, profanity, identity attack, insult, and threat. These attributes suggest the severity of toxic speech could vary from generic hateful comments (i.e., toxicity, severe toxicity, and profanity) to personalized attacks (i.e., identity attack or insult) and to the threat of causing physical harm (i.e., threats). While these forms of repression are known to be used against street protestors (Earl et al., 2022), how these manifest on social media and their effect on connective action is an interesting avenue to explore. Our research suggests toxicity is expressed through both text and images; however, image postings contain more toxicity than text postings. We also observed a positive correlation between toxicity attributes and interactions, especially for image postings. Additionally, toxicity attributes generally increased interactions with postings. However, toxicity attributes in text posting reduced the interactions compared to image postings. In other words, collective engagement increases for toxic image postings and decreases for toxic text postings.

The timing also determines the responses to (counter)connective repression. For both pro-choice and pro-life groups, toxic image postings increased interactions in the pre-leak period, whereas toxic text postings increased interactions in the post-fallout period. In the post-leak period, for pro-choice, the toxicity attributes in images had more correlation with interactions, and for pro-life, the toxicity attributes in text postings had more correlation with interactions. Again, for pro-choice, compared to images, the text postings with profanity had a negative effect on interactions. In the fallout period, for pro-choice, the correlation of toxicity attributes, especially profanity with interactions, is more for text postings. For pro-life, the correlation of toxicity attributes and interactions remains more for images, except for identity attack, which is stronger for text. In the post-fallout period, there was an increase in interactions for both groups. For pro-choice, the identity attack in text postings increased the interactions. For pro-life, the severe toxicity, profanity, and identity attack in text postings increased interactions.

To that end, this study contributes to connective action research in three ways. ***First***, we examined toxicity in social movements and contributed to the scholarship that has so far been focused on the "good" side of social media to generate and scale digital movements (Syed & Silva, 2023; Young et al., 2019). ***Second***, we contribute by providing insights into opposing movements. The existing literature has been focused on a single movement, and as a result, theoretical advancements have been limited to the use of social media to take issues forward. However, social movement literature provides several valuable elements to understand the emergence of competing movements, especially on social media, that engage in conscious and collective attempts to support or oppose changes (Shahin, 2023; Tarrow, 2011). We examined toxicity in the social media postings of pro-choice and pro-life and how it impacts collective engagement. This allowed us to forward a theory on repressive connective action, which opens an area for future research. ***Third***, we contribute by examining the content of both images and text postings and enhance the scholarship that so far paid little attention to images. As our results suggest, the content of the images can portray more toxicity than text postings. Toxic images also lead to increased interactions. Future research could further examine the engagement with toxic and hate speech portrayed in visual imagery.

The findings have implications for social media platforms and social movement activists. Social media platforms rely on a combination of artificial intelligence, user reporting, and content moderators to enforce their rules regarding appropriate content. If the engagement with the toxic posts increases, the algorithms driving the platforms are designed to maximize user engagement by recommending content with high interactions, which would inadvertently promote toxic content. The rules also vary by platform, and toxicity is more prevalent on some platforms (Amnesty International, 2023). Thus, platforms have a civic responsibility to implement techniques and processes that detect toxicity considered disrespectful in a broader context, which can lead to more sensible social discourse. In a similar vein, toxic discourse could have detrimental effects on

social movement organizers and activists. Instead of strengthening the voices of those who fight for justice and equality, the toxicity that activists, especially women and minorities, experience could lead them to self-censor what they post, limit their interaction, or even drive them off the platforms completely, pushing them back to a culture of silence (Amnesty International, 2018).

Finally, this study has some limitations that open further avenues for research possibilities. First, we focused on the Roe v. Wade movement, a highly controversial and divisive context, which might make toxicity more apparent. Future research will examine toxicity in other social movements where issues other than gender are salient. Second, our data collection is guided by a set of keywords or hashtags that may not entirely represent pro-choice or pro-life groups. For instance, while a hashtag might represent a specific group, the content of the posting might not. Thus, future research will consider more robust means to classify the posts and enhance the validity of the dataset. Additionally, the data from Instagram includes public posts from influential and verified users and does not represent the ideas of "ordinary" citizens. It would be interesting to examine the toxicity in "ordinary" citizens' postings and how that impacts collective engagement. Third, we relied on Google's API to compute the toxicity scores. While Google continues to improve the machine learning model to increase accuracy, the algorithm might miss or over-identify certain attributes (Hemphill, 2022). Finally, there is a scope to extend the analysis by examining the source (i.e., activists, general supporters, or opposers) and the target (i.e., activists or bystanders) of toxicity and thereby provide a more fine-grained topology of digital repression.

6. References

Amnesty International. (2018). *Toxic Twitter - A Toxic Place for Women*. https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1-1/

Amnesty International. (2023). *Hateful and abusive speech towards LGBTQ+ community surging on Twitter surging under Elon Musk*. https://www.amnesty.org/en/latest/news/2023/02/hateful-and-abusive-speech-towards-lgbtq-community-surging-on-twitter-surging-under-elon-musk/

Anti-Defamation League. (2021). *Online Hate and Harassment: The American Experience*. https://www.adl.org/resources/report/online-hate-and-harassment-american-experience

Ayoub, P. M., & Chetaille, A. (2020). Movement/countermovement interaction and instrumental framing in a multi-level world: Rooting Polish lesbian and gay activism. *Social Movement Studies*, *19*(1), 21-37.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, *15*(5), 739-768.

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge University Press.

Berente, N., Seidel, S., & Safadi, H. (2019). Research commentary—data-driven computationally intensive theory development. *Information Systems Research*, *30*(1), 50-64.

Bharati, P., Lee, C., & Syed, R. (2019). Influence of Trolling on Social Media Participation: An Empirical Investigation. Proceedings of the Academy of Management.

Brenan, M. (2021). *Record-High 47% in U.S. Think Abortion Is Morally Acceptable*. https://news.gallup.com/poll/350756/record-high-think-abortion-morally-acceptable.aspx

Chang, R.-C., Rao, A., Zhong, Q., Wojcieszak, M., & Lerman, K. (2023). # RoeOverturned: Twitter Dataset on the Abortion Rights Controversy. *arXiv preprint arXiv:2302.01439*.

CrowdTangle Team. (2023). *CrowdTangle. Facebook, Menlo Park, California, United States*. https://crowdtangle.com/

Delaney, N. (2022). *Roe v. Wade has been overturned. What does that mean for America?* https://www.hks.harvard.edu/faculty-research/policy-topics/fairness-justice/roe-v-wade-has-been-overturned-what-does-mean#erica-chenoweth

Earl, J., & Beyer, J. L. (2014). The dynamics of backlash online: Anonymous and the battle for WikiLeaks. In *Intersectionality and social change*. Emerald Group Publishing Limited.

Earl, J., Maher, T. V., & Pan, J. (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances*, *8*(10), eabl8198.

Edwards, C. (2022). *Influencers and social media react to the Roe v Wade ruling*. https://corq.studio/insights/influencers-and-social-media-react-to-the-roe-v-wade-ruling/

Fixmer-Oraiz, N. (2022). No going back: The struggle for a post-Roe reproductive justice. *Quarterly Journal of Speech*, *108*(4), 426-430.

Gallagher, R. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2018). Divergent discourse between protests and counter-protests:# BlackLivesMatter and# AllLivesMatter. *PloS One*, *13*(4), e0195644.

George, J. J., & Leidner, D. E. (2019). From clicktivism to hacktivism: Understanding digital activism. *Information and Organization*, *29*(3).

Google. *Perspective API*. https://www.perspectiveapi.com/

Halfmann, D., & Young, M. (2010). War pictures: The grotesque as a mobilizing tactic. *Mobilization: An International Quarterly*, *15*(1), 1-24.

Hemphill, L. (2022). *What Social Media Platforms Miss About White Supremacist Speech*. https://www.adl.org/resources/report/very-fine-people

Inata, K. (2021). Protest, counter-protest and organizational diversification of protest groups. *Conflict Management and Peace Science*, *38*(4), 434-456.

Lantos, H., Pliskin, E., Wildsmith, E., & Manlove, J. (2022). *State-level Abortion Restrictions Will Negatively Impact Teens and Children*. https://www.childtrends.org/blog/state-level-abortion-restrictions-will-negatively-impact-teens-and-children

Leong, C., Faik, I., Tan, F. T., Tan, B., & Khoo, Y. H. (2020). Digital organizing of a global social movement: From connective to collective action. *Information and Organization*, *30*(4).

Leong, C., Pan, S. L., Bahri, S., & Fauzi, A. (2019). Social media empowerment in social movements: power activation and power accrual in digital activism. *European Journal of Information Systems*, *28*(2), 173-204.

Massanari, A. (2017). # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, *19*(3), 329-346.

McGarty, C., Haslam, S. A., Turner, J. C., & Oakes, P. J. (1993). Illusory correlation as accentuation of actual intercategory difference: Evidence for the effect with minimal stimulus information. *European Journal of Social Psychology*, *23*(4), 391-410.

McKenna, B. (2020). Creating convivial affordances: A study of virtual world social movements. *Information Systems Journal*, *30*(1), 185-214.

Meyer, D. S., & Staggenborg, S. (1996). Movements, countermovements, and the structure of political opportunity. *American Journal of Sociology*, *101*(6), 1628-1660.

Miranda, S., Berente, N., Seidel, S., Safadi, H., & Burton-Jones, A. (2022). Editor's comments: computationally intensive theory construction: a primer for authors and reviewers. *MIS Quarterly*, *46*(2), iii-xviii.

Miranda, S. M., Young, A., & Yetgin, E. (2016). Are social media emancipatory or hegemonic? Societal effects of mass media digitization. *MIS Quarterly*, *40*(2), 303-329.

New York Times. (2022). *Supreme Court Rules on Abortion: Thousands Protest End of Constitutional Right to Abortion*. https://www.nytimes.com/live/2022/06/24/us/roe-wade-abortion-supreme-court

Nurik, C. L. (2022). Facebook and the Surveillance Assemblage: Policing Black Lives Matter Activists & Suppressing Dissent. *Surveillance & Society*, *20*(1), 30-46.

Oh, O., Eom, C., & Rao, H. R. (2015). Research note—Role of social Media in Social Change: An analysis of collective sense making during the 2011 Egypt revolution. *Information Systems Research*, *26*(1), 210-223.

Pan, J., & Siegel, A. A. (2020). How Saudi crackdowns fail to silence online dissent. *American Political Science Review*, *114*(1), 109-125.

Park, J. H., Moon, J. Y., & Hong, S.-J. (2021). Understanding the bi-directional message diffusion mechanism in the context of IT trends and current social issues. *Information & Management*, *58*(8).

Rohlinger, D. A., & Klein, J. (2012). Visual landscapes and the abortion issue. *American Behavioral Scientist*, *56*(2), 172-188.

Ross, L. (2019). I'm a black feminist. I think call-out culture is toxic. *The New York Times*.

Selander, L., & Jarvenpaa, S. L. (2016). Digital action repertoires and transforming a social movement organization. *MIS Quarterly*, *40*(2), 331-352.

Shahin, S. (2023). Affective polarization of a protest and a counterprotest: Million MAGA March v. Million Moron March. *American Behavioral Scientist*, *67*(6), 735-756.

Sutton, B., & Vacarezza, N. L. (2020). Abortion rights in images: Visual interventions by activist organizations in Argentina. *Signs: Journal of Women in Culture and Society*, *45*(3), 731-757.

Syed, R., & Silva, L. (2023). Social Movement Sustainability on Social Media: An Analysis of the Women's March Movement on Twitter. *Journal of the Association for Information Systems*, *24*(1), 249-293.

Tarafdar, M., & Ray, K. D. (2021). Role of Social Media in Social Protest Cycles: A Sociomaterial Examination. *Information Systems Research*, *32*(3), 675-1097.

Tarrow, S. G. (2011). *Power in movement: Social movements and contentious politics*. Cambridge University Press.

Tim, Y., Pan, S. L., Bahri, S., & Fauzi, A. (2018). Digitally enabled affordances for community-driven environmental movement in rural Malaysia. *Information Systems Journal*, *28*(1), 48-75.

Tye, M., Leong, C., Tan, F., Tan, B., & Khoo, Y. H. (2018). Social media for empowerment in social movements: the case of Malaysia's grassroots activism. *Communications of the Association for Information Systems*, *42*(1), 15.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social Media Affordances for Connective Action: An Examination of Microblogging Use During the Gulf of Mexico Oil Spill. *MIS Quarterly*, *41*(4), 1179-1205.

Weidmann, N. B., & Rød, E. G. (2019). *The Internet and political protest in autocracies*. Oxford Studies in Digital Poli.

Young, A., Selander, L., & Vaast, E. (2019). Digital organizing for social impact: Current insights and future research avenues on collective action, social movements, and digital technologies. *Information and Organization*, *29*(3).

Zald, M. N., & Useem, B. (1987). Movement and countermovement interaction: Mobilization, tactics, and state involvement. In M. N. Zald & J. D. McCarthy (Eds.), *Social Movements in an Organizational Society* (pp. 247-272). Transaction Publishers.