

# The Impact of Culture on Online Toxic Disinhibition: Trolling in India and the USA

Pnina Fichman  
Rob Kling Center for Social Informatics,  
Indiana University  
[fichman@iu.edu](mailto:fichman@iu.edu)

Maanvi Rathi  
Rob Kling Center for Social Informatics,  
Indiana University  
[mrathi@iu.edu](mailto:mrathi@iu.edu)

## Abstract

*The pervasiveness of online trolling has been attributed to the effect of online toxic disinhibition, suggesting that perpetrators behave in less socially desirable ways online than they do offline. It is possible that this disinhibition effect allows for everyone to start on a level playing field online, regardless of race, gender, or nationality, but it is likewise possible that the disinhibition effect is context-dependent and sensitive to socio-cultural variations. We aim to explore if toxic online disinhibition effects depend on national culture and gender by examining the extent of trolling towards tweets by Americans and Indians, from both genders. Content analysis of 3,000 Twitter posts reveals that significantly more trolling comments were posted on tweets by Americans than by Indians, and on tweets by women than men. We conclude that the online disinhibition effect may exacerbate, replicate, or mediate existing socio-cultural differences, but it does not eliminate them.*

## 1. Introduction

With fewer global barriers to access and use of social media platforms, trolling is now everywhere. More than 500 million tweets are posted daily, spreading misinformation, disinformation and trolling, leading Twitter to announce new ways to combat the latter [1]. These trolling behaviors, like other less socially desirable activities, are largely enabled by the toxic online disinhibition effect [2]. Online, people feel less visible, more anonymous, and much of their interactions with others are asynchronous, as a result their sense of power dynamics is distorted, and they are more likely to disassociate themselves from their online actions once they are offline. This leads to an increase in online behaviors that are not regulated in the same way as real

life; then, more trolling, harassment and abuse are noticeable on online platforms. As trolling behaviors spread widely, research follows with significant efforts to detect trolling, identify trolls' motivations, behaviors, and tactics, to understand the impact of trolling on individuals and communities, and unpack perceptions of and reactions to trolling [e.g., 3, 4, 5, 6, 7, 8]. What becomes clear is that online trolling is multidimensional and context-dependent, thus varying from one setting to another [8]. Whereas less socially desirable behaviors more prominently surface online, it is possible that trolling may exacerbate, replicate, or mediate existing socio-cultural differences. It is also possible that certain attributes of trolling targets play a role in trolling as well: "harassment on the basis of race and gender is indicative of trolling" [7, p. 1]; yet, research mostly focuses on online perpetrators. Early research on the impact of trolling targets on the extent and type of trolling, reveals that not only the gender of the perpetrator, but also the gender of target plays a role in online trolling [e.g., 9, 10]. Individuals from different genders troll differently [e.g., 11], and women have been the target of more online trolling than men [9].

Other attributes of trolling targets and perpetrators may have an impact on the extent and type of trolling as well. For example, it is possible that trolling targets in different countries will have different experiences. The theory of cultural relativity suggests that different cultures have their own standards of acceptable behaviors [10]. Online disinhibition is partially enabled by reduced status, power, and authority, compared to face-to-face [2], and since status and power distance vary between countries [12], national culture of trolling targets cannot be ignored. Power differences online may be as prominent or even more noticeable than they are offline, and trolling targets in high power distance cultures (India, for example) may have different trolling experiences than those in low power distance cultures (USA, for example). However, while power dynamics

may shift online [2], power differences may become unnoticeable and as such, regardless of demographics [10] powerless or powerful trolling targets may be treated equally bad. It is also possible that those in position of political power will be the target of more trolling online than those who are less powerful. Nonetheless, there is little research on the extent to which targets' national culture have on their trolling experience. Our study is informed by and aims to contribute to Suler's (2004) disinhibition effect theory, by addressing the following research question: Does trolling vary based on target's national culture (country of residency)? We report significant differences in the scope and type of trolling between Twitter users in the USA and India. We provide socio-cultural explanations to these differences and suggest future research directions. Our study makes a significant contribution to the theory of online disinhibition, with empirical evidence, showing that national culture plays a role in trolling; this means that the online toxic disinhibition effect does not simply equalize the playing field for all users online, but that instead it is sensitive to cultural variations. We conclude arguing that the online toxic disinhibition is context dependent and varies from one setting to another.

## 2. Theoretical Background

### 2.1 The Disinhibition Effect

People behave differently online than they do offline; they say and do things that are less socially acceptable and desirable. Some simply loosen up and more openly reveal themselves to others, while those who feel less restrained act in hurtful ways, harming individuals and groups. This phenomenon has been termed the online disinhibition effect, and it has two opposing manifestations. The first, benign disinhibition, describes the act of revealing more emotions and kindness; and the second, toxic disinhibition, describes the act of using rude language, or revealing hate, anger, and strong criticism of others. Online trolling is enabled by the toxic online disinhibition effect. Prior research focused attention on the causes for toxic online disinhibition more than benign disinhibition [13].

Suler (2004) suggests that the disinhibition effect is caused by six factors: 1) dissociative anonymity: when people can hide their online identity and behave in less socially desired ways, even disassociating their online selves from their real offline selves; 2) invisibility: people don't see or hear each other on online textual platforms; the invisibility gives way to a comfort which enables them to do things they otherwise would not; 3) asynchronicity: when people don't have to worry about others' immediate reactions to their behaviors and

words on asynchronous platforms, disinhibition is increased; 4) solipsistic introjection: as the boundaries of self and others are blurred in the absent of face-to-face cues, a person assigns a voice and face to the words of their online companion, shaped by the person's expectation in addition to their companion's words; 5) dissociative imagination: the combination of the above factors result in the belief that when people turn off the computer, their actions and imagined companion are simply left behind and have nothing to do with their offline reality; and 6) minimization of authority: when online, offline cues of authority (dress, body language) disappear; the elimination of elevated positions results in the unanimous feeling of possessing an equal voice. Thus, "[e]veryone—regardless of status, wealth, race, or gender—starts off on a level playing field." [2, p. 324].

We aim to gain a nuanced understanding of this sixth factor in Suler's framework and its manifestation in online trolling on Twitter, because power dynamic still play an important role in online trolling. For example, celebrities have more power and influence than others in society and online equalizations bring celebrities closer to everyone, becoming then more accessible to all online users. At the same time, trolls can troll all online users equally; instead, they can choose who to troll, based on factors like their target's gender, race, or nationality. When trolling women more, trolls reinforce existing power imbalance also online. Furthermore, cultural differences in power dynamics that exist offline within a culture, may impact trolling online, impacting the online toxic disinhibition effect as well. We focus on the sixth factor in Suler's framework to see if the level playing field online is indeed equal or if existing offline differences (e.g., power imbalance) still matter when it comes to the online toxic disinhibition effect, and the extent of trolling different target groups are susceptible to receive. There is still a need to understand the impact of national culture on online disinhibition and the variations of the effect between countries. It is possible that cultural norms, which regulate behaviors offline, may impact toxic online disinhibition effect as well.

### 2.2 Hofstede's Cultural Dimensions

Hofstede's cultural dimensions framework [12] has been treated as a paradigm in cross cultural studies. While this framework has received a lot of criticism, it is perhaps the most widely used framework in cross-cultural research. It includes six cultural values: power distance, individualism-collectivism, masculinity-femininity, uncertainty avoidance, long term-short term orientation, and indulgence-restraint [12].

Power distance is defined as "the extent to which the less powerful members of organizations and

institutions (like the family) accept and expect that power is distributed unequally" [12, p. 80]. Individualism refers to a society in which "ties between individuals are loose: everyone is expected to look after himself or herself...[while] Collectivism... pertains to societies in which people...are integrated into strong, cohesive in groups" [p. 114]. Masculinity relates to societies in which social "gender roles are clearly distinct: men are supposed to be assertive, tough...whereas women are supposed to be modest, tender, and concerned with quality of life" [p. 170]; "femininity pertains to societies in which social gender roles overlap [p. 171]. Uncertainty avoidance is defined as "the extent to which the members of a culture feel threatened by ambiguous or unknown situations" [p. 229]. Long term orientation refers to societies that are "oriented towards future reward... perseverance, and thrift, [while] short term orientation... [refers to societies that are] oriented towards past and present...respect of tradition, preservation of "face," and fulfilling social obligations" [pp. 284-285]. Indulgence stands for a society that "allows relatively free gratification of basic and natural human drives related to enjoying life and having fun...[whereas] restraint societies...[believe] that such gratification needs to be curbed and regulated by strict social norms" [p. 333].

In our study we focus on two countries, the USA and India, representing high and low power distance ranks; we have used country of residency of the target account holders. Table 1 provides the scores and ranks for the USA and India, from hofstede-insights.com, on each of these six cultural dimensions, as well as the range of scores and the number of countries ranked on each dimension.

**Table 1. Country comparison on six Hofstede's dimensions**

Dimension	India score	USA score	Range of scores
Power distance	77	40	11-104
Masculinity	56	62	5-110
Long term	51	26	0-100
Individualism	48	91	6-91
Uncertainty avoidance	40	46	8-112
Indulgence	26	68	0-100

### 3. Relevant Work

Just a few years ago, more than 340 million Twitter users post over 500 million tweets every single day, and now only about 199 million active users tweet daily; while one of five American adults is an active Twitter

user, most of Twitter's users (262 million) are not American [14]. But with this social media surge also comes a steady growth of abuse, misinformation, trolling and harassment [15, 16]. There are many users who "exploit the platform to promote a hostile environment and spread false information and propaganda," even as most users post content related to their "innocuous feelings, opinions, ideas, and information" [16, p. 152]. The online disinhibition effect, along with other sociotechnical affordances, enables these anti-social behaviors on social media platforms. Yet, Twitter is perhaps more susceptible to public debate and trolling due to its enabling architecture: users are granted the ability to write short tweets along with the creative use of memes [17]. They can also be more "linguistically and argumentatively simple" due to character restrictions, more "impulsive" because of the convenience of posting on the go with a device like a mobile phone, and more "uncivil due to informality and depersonalized interactions" [18, p. 60]. With Twitter's anonymity and reach, harassment and hate speech have become part of a "landscape that is abstract and beyond the realms of traditional law enforcement" [19, p. 83]. As such, Twitter is one of the main hubs of social media trolling; what leads to more trolling on the platform is the fact that many public figures and politicians are regular users of Twitter.

Because political trolling is widespread on the platform, recent scholarly attention was given to the automatic identification of online trolling on Twitter through linguistic analysis of trolling datasets [e.g., 5, 20, 21, 22, 23]. Efforts to identify trolling on Twitter are important, as these trolls negatively affect the political process, cause distrust in the political systems, and increase political polarization and conflict [24]. Most of these efforts aim to identify troll accounts by Russia's Internet Research Agency (IRA), which attempted to interfere with the 2016 U.S. election; the IRA utilized fake accounts on Twitter, known as "Russian trolls," that aimed at spreading misinformation and politically biased information. The "Kremlin trolls" simply aim to undermine American political institutions, spread discord within American society, and undermine the United States' global influence [25]. Scholars [23] focused on anti-democratic discourse and investigated the linguistic profile of Twitter trolls and found that trolls tend to write shorter posts and used a smaller number of lexical types and tokens. Others [22] developed machine learning models that predict whether a Twitter account is a Russian troll, using both behavioral and linguistic features. They found that Russian troll accounts on Twitter were not merely bots, but instead managed their online identities in various complex ways. Another study [20] examined how trolls attempted to manipulate public opinion; they identified

theoretically grounded linguistic markers of deception and showed that deceptive language cues can help identify trolls accurately. While these studies are important, they largely focused on the trolls and disregard attributes of trolling targets.

Political trolling is spreading also around the globe as politicians use social media platforms more often to promote their political agendas [e.g., 17, 26, 27, 28, 29]; yet little is known about the variations of trolling across countries [30]. According to Evolvi [31], 31% of tweets referred to political leaders like Donald Trump, Sadiq Khan, Barack Obama, Hillary Clinton, and Angela Merkel, indicating that “tweets discuss politics globally,” alluding to both left- and right-wing politics [17, p. 7]. The “trolls’ political performance is a global phenomenon” [17, p. 4097], and many instances of state-sponsored trolling within and across countries are driven by political agendas [e.g., 32]. There are growing trends of political trolling in Turkey, Russia, India, UK, and the USA [17]. Satirical political trolling manifested cultural differences and was more prevalent in countries with low power distance than high power distance scores [30].

It is possible that tweets from certain countries may attract different amounts of trolling. Using Hofstede’s power distance dimensions [12], we examine whether trolling towards tweets by individuals from high power distance (India) and low power distance (USA) countries differ. Because India is ranked much higher than the USA on the power distance index (Table 1), we expect more trolling to happen towards American Twitter users than Indian Twitter users, and propose to test the following hypothesis:

H1: Tweets by Americans will be the targets of more trolling than tweets by Indians.

This might be exacerbated by other attributes of trolling targets, specifically by gender. According to research done on online trolling of UK Members of Parliament, female politicians were subject to more racial and sexual abuse—along with misogynistic comments—than men, who were attacked based on their professional duties [15]. In the same study, males reported concern on their professional damage whereas females expressed concern about personal safety; however, females faced more emotional and psychological stress, in addition to overall damage to their reputation. In another study, tweets by female politicians were the target of more trolling than tweets by male politicians [9]. In a study of sex differences in anti-social behaviors, men reported more antisocial motives for using Facebook than did women, while women reported stronger prosocial motives for using Facebook [33]. Based on another analysis of Twitter

trolls targeting three female journalists in India, Regu [34, p. 437] concluded that “social media platforms constitute convenient havens of harassment against assertive women.” Ortiz [7, p. 6] claimed that “trolling is a mechanism through which White men, especially politically conservative men, collectively target others with their rage, disgust, and discontent”; while this claim might be too extreme, trolls are motivated at times to troll because of their target’s gender. Trolling has been widely associated with males as trolls, and females as the target of trolling [e.g., 9, 34, 10], and cyber-harassment typically targets females more frequently than males [11]. Scholars [11, p. 141] call for research that looks at the intersection of gender with “other aspects of an individual’s identity in their experiences of and vulnerabilities to cyberviolence,” including trolling, and thus we propose to test the following hypothesis:

H2: Tweets by female Americans will be the targets of more trolling than tweets by others.

Using Twitter data, we test for the impact of the target’s national culture (country of residency: USA/India) and gender (female/male), as well as the impact of the intersection of the two on the scope of trolling.

#### 4. Method

We collected data from Twitter on February 9-23, 2020, using ExportComments.com, and utilizing a Twitter account that we created for this purpose. We collected 200 tweets from eight Twitter accounts of celebrities from two countries (Table 2) (most recent 25 tweets per account) and all their respective comments. To reduce confounding variables effect, these eight celebrities had either similar stature, social media presence, or popularity in the respective countries. Then, we randomly sampled 15 comments per tweet, which resulted in 375 comments per account and a total of 3,000 comments in our sample. Our sample included an equal number of comments per account, and an equal number of comments by country of residency and by gender.

**Table 2. Gender and country of residence per account**

Account Holder	Country of Residence	Gender
Abhishek Bachchan	India	Male
Alexandria Ocasio-Cortez (AOC)	USA	Female
Anushka Sharma	India	Female
Donald Trump	USA	Male

Kevin Spacey	USA	Male
Lindsay Lohan	USA	Female
Mamata Banerjee	India	Female
Narendra Modi	India	Male

We uploaded the comments into NVivo12, a software for qualitative data analysis, to facilitate content analysis. Then we developed a codebook with eight trolling tactics through an iterative process of coding and discussion among the authors, and tested it on a sample of our data (Table 3). We then coded the data at the individual post level; each of the 3,000 posts was coded with all the relevant codes. Inter-coder reliability was 85% simple agreement.

## 5. Findings

**Table 3. Trolling tactics and their frequencies**

Tactic	Definition	Total	India	USA
Derailment	Purposefully leading a conversation off track.	354	167	187
Hyperbole	Exaggerating one's strengths or another's weaknesses.	87	20	67
Ideological misalignment	Comments made because of a difference in political opinions.	152	23	129
Ideologically extremizing language	Contains extremist language used to critically describe a subject/ their behavior.	96	16	80
Insulting	Statement meant to insult an individual or group of people. Includes swearing, mocking and name-calling.	197	40	157
Personal attacks	Statement meant to target an individual. Includes character	195	42	153

	assassination			
Provocation	Statement intended to elicit a specific reaction.	284	63	221
Pseudo sincere	Providing the impression of particular motivations while actually being driven by other motivations.	346	95	251
Total		1,711	466	1245

We found that some trolling behaviors were more common than others, with derailment, provocation, pseudo-sincere, personal attacks, and insulting as the most frequent trolling techniques (Table 3). Trolling frequencies varied by country of residence. We found more trolling comments in the USA than in India, and more trolling of females' tweets than of males' (Table 4). We tested for differences across target groups using cross-tabulation statistics and found that these differences were statistically significant (Table 5), meaning that both country of residency and gender of trolling targets has an impact on the extent of trolling targets receive; the combination of the two (gender by country of residency) further exacerbate this impact. Thus, we found support for both of our hypotheses (Table 6).

**Table 4. Frequency of trolling by country of residence and gender**

Demographics	Female	Male	Total
India	211	255	466
USA	699	546	1,245
Total	910	801	1,711

**Table 5. Results of statistical tests of differences across categories**

Demographics	$\chi^2$
Gender	16.161***
Country of residence	825.454***
Gender by country	16.078***

\* = p < .05, \*\* = p < .01, \*\*\* = p < .001.

**Table 6. Results of hypotheses testing**

Hypothesis	Results
H1: Tweets by Americans will be the targets of more trolling than tweets by Indians.	Supported
H2: Tweets by female Americans will be the targets of more trolling than tweets by others.	Supported

## 6. Discussion

Our study makes two major contributions that will be the focus of our discussion. First, our findings show that not all trolling targets are created equal. Trolling targets experience different levels and types of trolling, depending on the country they live in and on their gender. Second, cultural diversity plays a role in toxic online disinhibition. In contradiction to Suler's (2004) suggestions that users, regardless of status, wealth, gender, and race are positioned online equally. Our study suggests that these differences are not eliminated online. We argue that the online disinhibition effect is sensitive to target attributes.

First, our study expands prior research that suggests that the extent and nature of trolling is context dependent and varies between countries; we demonstrate that tweets by American celebrities are subject to more trolling than tweets by Indian celebrities. The variations might be due to socio-cultural differences, as there are significant differences between the two nations' cultural values [12, 35]. The USA is a society with low power distance "which might provide a better environment for trolling" [30, p. 10] than India. In fact, the two countries are ranked differently on all of Hofstede's dimensions (Table 1), most notably when it comes to power distance, individualism, and indulgence; cultural norms in the USA may be more accepting of trolling and toxic disinhibition.

It is possible that the differences in trolling targets from the two countries are due to the fact that India is a more authoritarian state, where trolling may occur less due to the perceived lack of online anonymity, that results in higher probability of censorship or self-censorship and in fear of consequences in real-life [30]. The lack of perceived anonymity in India, which is detrimental to online disinhibition [2], limits the extent of trolling in that country.

Moreover, censorship may be higher in India than in the USA, despite the fact that the value of free speech is part of both nations' constitutions. In the USA, the First Amendment ensures the freedom of speech for all citizens of America and restricts any prohibition of the same. In India, Article 19(1)(a) of the Indian Constitution also grants freedom of speech and

expression to all its citizens. While both constitutions allude to the importance of free speech, it has been practiced differently in the two countries, perhaps due to differences between their cultures. The USA is a universalistic culture and India is a particularistic culture [35]. In universalistic cultures, rules are more important than relationships, and thus the public adhere to the rules, while in particularistic cultures, relationships and other situational factors are more important than rules, which thus serve only as a recommendation, a point of departure [35]. Because India is a particularistic culture [36], Indians may perceive Article 19(1)(a) of their constitution only as a recommendation and not a promotion of free speech. As a result, trolling is less common in India than in the USA.

Still, it is possible that the magnitude of trolling in the two countries differ simply because of differences in social media use and internet availability among the masses. Even though India accounts for more internet users and social media users than the USA, the USA accounted for almost 4.5 times more Twitter users than India in 2020. Future research could examine trolling variations between countries with similar social media use patterns, but different cultures, in order to test if and to what extent cultural factors play a role in explaining trolling variations.

Second, our study provides a nuanced understanding of the disinhibition effect, showing that national culture matters; online disinhibition may exacerbate, replicate, or mediate existing socio-cultural biases, but it does not eliminate them. Cultural norms regulate behaviors online and offline, and the online disinhibition effect does not eliminate them. As power dynamics shift online [2], trolling behaviors are not equally distributed across targets demographics but differ between genders and across cultures, despite minimization of authority and power dynamics online. Sociocultural biases persist online and the disinhibition effect may only intensify their manifestation, at least in terms of gender and country of residency. Indeed, the online disinhibition effect does not eliminate existing socio-cultural biases and service discrimination persists online [37]. While service discrimination may not be intentional, trolling is.

Online disinhibition is partially enabled by reduced status, power, and authority online, but power distance varies between countries [12], and national cultural norms seem to play a dominant role in trolling, exacerbating other socio-cultural variations and impacting trolling. As a result, the online disinhibition effect may not be as prominent in some cultures as in others, as cultures have their own standards for acceptable and deviant behaviors [10]. Future research

may examine the more nuanced cultural disparities in the effect of online disinhibition.

In our study, targets' genders were not hidden from Twitter users, and while disinhibition may cause perpetrators to behave badly towards all users regardless of their gender, we found that trolls target females more often. Suler and Phillips [10, p. 275] claim that there are more male perpetrators online than female, and that "online males...have a more difficult time restraining...their Eros-ridden nature...lacking in the compassion...that is needed to realize how other users are...real people. Maybe there simply are more male users...resulting in a "boys club" atmosphere that encourages...the harassment of women." While this last argument is in line with our findings (more trolling towards females than males), it is in sharp contrast to Suler's (2004) argument that everyone, regardless of gender, for example, starts off equally online. Clearly, when studies show that female targets are trolled more frequently than males, it is impossible to make the case that females and males are treated equally online, at least when it comes to trolling on Twitter. Future research may examine whether targets' genders matter across platforms. Perhaps there will be less of a disparity between trolled genders on platforms where the majority of users are women. Thus, we conclude that sociocultural factors, such as the demographics of trolling targets, are impactful online, suggesting a nuanced disinhibition effect. Future research may examine if socio-cultural differences online exacerbate, replicate, or mediate existing socio-cultural biases.

## 7. Limitations

One of the limitations of the study is that our sample of accounts may allow for intervening variables to mitigate our findings. For example, it is possible that in addition to the targets' demographics, targets' personality and political views may play a role in the extent and type of trolling their tweets receive. Another limitation is that we chose only two countries to examine the impact of the targets' countries of residency on trolling, yet trolling in other countries may differ. In addition, online and social media platforms, other than Twitter, could be hubs for social media trolling. Studying these communities may explain other nuances of the socio-cultural impact of social media trolling. Still, our comparison, even if limited in scope of countries, is valuable as it provides the first step toward future cross-cultural research of the effect of online disinhibition that may include more countries.

## 8. Conclusions

The major contribution of the paper is our demonstration that the online toxic disinhibition effect is not blind to demographics of trolling targets. We demonstrate that the level of trolling towards users' tweets vary based on the target's country of residence and gender. Thus, we contribute to Suler's (2004) online disinhibition framework by showing that the disinhibition effect does not eliminate existing socio-cultural biases, but rather exacerbate, mediate, or alleviate them; targets' demographics matter. Specifically, "minimization of authority" online [2], does not completely equalize users on Twitter.

This is important because prior research largely ignored cross-cultural variations in online disinhibition. Our study highlights the need for future research on 1) trolling in different countries; 2) the extent that toxic disinhibition effect mediates, enhances, or reduces existing sociocultural biases online; 3) the degree to which cultural norms at different levels (nations, professions, communities, and small groups) may mediate each other in facilitating online trolling; 4) trolling on other social media and online platforms.

## 9. References

- [1] J Brandon, "Twitter just announced a new way to combat trolls, but only if they are stupid", *Forbes*, retrieved from <https://www.forbes.com/sites/johnbbrandon/2020/01/09/twitter-just-announced-a-new-way-to-combat-trolls-but-only-if-they-are-stupid/#2560168c4258>, 2020, January 9.
- [2] J. Suler, The online disinhibition effect, *Cyberpsychology & Behavior*, 7, 2004, 321–326.
- [3] C. Cook, J. Schaafsma, and M. Antheunis, Under the bridge: An in-depth examination of online trolling in the gaming context, *New Media & Society*, 20(9), 2018, 3323–3340. <https://doi-org.proxyiub.uits.iu.edu/10.1177/146144481774857>
- [4] I. Guy, and B. Shapira, From royals to vegans: Characterizing question trolling on a community question answering website, In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*, ACM, New York, NY, 2018.
- [5] C. Hardaker, Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions, *Journal of Politeness Research*, 6(2), 2010, 215–242. <https://doi-org.proxyiub.uits.iu.edu/10.1515/JPLR.2010.011>
- [6] S. Monakov, Early detection of internet trolls: Introducing an algorithm based on word pairs / single words multiple repetition ratio, *PLoS ONE* 15(8): e0236832, 2020. <https://doi.org/10.1371/journal.pone.0236832>
- [7] S. M. Ortiz, Trolling as a collective form of harassment: An inductive study of how online users understand trolling. *Social Media & Society* (first published June 23,

- 2020). <https://doi.org/proxyiub.uits.uu.edu/10.1177/2056305120928512>
- [8] M. R. Sanfilippo, P. Fichman, and S. Yang, Multidimensionality of online trolling behaviors, *The Information Society*, 34(1), 2018, 27–39.
- [9] P. Fichman, M. W. McClelland, The impact of gender and political affiliation on trolling, *First Monday*, 26(1), 2021. <https://firstmonday.org/ojs/index.php/fm/article/download/11061/10034>
- [10] J. Suler, and W. L. Phillips, The bad boys of cyberspace: Deviant behavior in a multimedia chat community, *CyberPsychology & Behavior*, 1, 1998, 275–294.
- [11] E. L. Backe, P. Lilleston, and J. McCleary-Sills, Networked individuals, gendered violence: A literature review of cyberviolence, *Violence and Gender*, 5 (3), 2018, 135-146.
- [12] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and organizations: Software of the mind*, 3rd Edition, New York: McGraw-Hill, 2010.
- [13] J. Stuart, and R. Scott, The Measure of Online Disinhibition (MOD): Assessing perceptions of reductions in restraint in the online environment, *Computers in Human Behavior*, 2021, 114. <https://doi.org/10.1016/j.chb.2020.106534>.
- [14] Omnicore, Twitter by the numbers, August 11, 2021, <https://www.omnicoreagency.com/twitter-statistics/>
- [15] S. Akhtar, and C. M. Morrison, The prevalence and impact of online trolling of UK members of parliament, *Computers in Human Behavior*, 99, 2019, 322–327.
- [16] I. Clarke, Stylistic variation in Twitter trolling. In: J. Golbeck (ed.), *Online Harassment, Human-Computer Interaction Series*, Springer, Cham, 2018, pp. 151-178.
- [17] E. Bulut, and E. Yörük, Digital populism: Trolls and political polarization of twitter in turkey. *International Journal of Communication*, 11, 2017, 25. <https://ijoc.org/index.php/ijoc/article/view/6702/2158>
- [18] I. Clarke, Functional linguistic variation in Twitter trolling, *International Journal of Speech, Language & the Law*, 26(1), 2019, 57-84.
- [19] C. Hardaker, and M. McGlashan, “Real men don’t hate women”: Twitter rape threats and group identity, *Journal of Pragmatics*, 91, 2016, 80–93.
- [20] A. Addawood, A. Badawy, K. Lerman, and E. Ferrara, Linguistic Cues to Deception: Identifying Political Trolls on Social Media, *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)*, 2019, 15-25.
- [21] S. Alhazbi, Behavior-based machine learning approaches to identify state-sponsored trolls on twitter, in *IEEE Access*, 8, 2020, pp. 195132-195141. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9239285>.
- [22] J. Im, E. Chandrasekharan, J. Sargent, P. Lighthammer, T. Denby, A. Bhargava, L. Hemphill, D. Jurgen, and E. Gilbert, Still Out There: Modeling and Identifying Russian Troll Accounts on Twitter. *WebSci '20: 12th ACM Conference on Web Science*, 2020, 1–10. <https://doi.org/10.1145/3394231.3397889>
- [23] J. Lundberg, and M. Laitinen, Twitter trolls: A linguistic profile of anti-democratic discourse. *Language Sciences* (Oxford), 79, 2020, 1-14. <https://doi.org/10.1016/j.langsci.2019.101268>
- [24] S. Bauman, *Political cyberbullying: Perpetrators and targets of a new digital aggression*, Praeger; Santa Barbara, CA, 2020.
- [25] A. Entous, E. Nakashima, and G. Jaffe, Kremlin trolls burned across the Internet as Washington debated options, *The Washington Post*, 2017, December 25. [http://www.washingtonpost.com/world/national-security/kremlin-trolls-burned-across-the-internet-as-washington-debated-options/2017/12/23/e7b9dc92-e403-11e7-ab50-621fe0588340\\_story.html](http://www.washingtonpost.com/world/national-security/kremlin-trolls-burned-across-the-internet-as-washington-debated-options/2017/12/23/e7b9dc92-e403-11e7-ab50-621fe0588340_story.html)
- [26] O. Z. Gökçe, E. Hatipoğlu, G. Göktürk, B. Luetgert, and Y. Saygin, Twitter and politics: Identifying Turkish opinion leaders in new social media, *Turkish Studies*, 15 (4), 2014, 671-688.
- [27] J. H. Parmelee, and S. L. Bichard, *Politics and the Twitter revolution: How tweets influence the relationship between political leaders and the public*, Lanham, MD: Lexington Books, 2012.
- [28] H. Rajput, Social media and politics in India: A study on Twitter usage among Indian political leaders, *Asian Journal of Multidisciplinary Studies* 2 (1), 2014, 63-69.
- [29] T. A. Small, What the hashtag? A content analysis of Canadian politics on Twitter, *Information, Communication and Society*, 14, 2011, 872-895.
- [30] P. Fichman, The role of culture and collective intelligence in online global trolling: The case of trolling trump’s inauguration speech. *Information, Communication, and Society*, 2020 <https://www.tandfonline.com/doi/abs/10.1080/1369118X.2020.1824006>
- [31] G. Evolvi, Hate in a tweet: Exploring internet-based Islamophobic discourses, *Religions*, 9(10), 2018, 307.
- [32] C. Nyst, and N. Monaco, *State sponsored trolling*, Palo Alto, CA: Institute for the Future, 2018, [https://www.iftf.org/fileadmin/user\\_upload/images/Digital/IFTF\\_State\\_sponsored\\_trolling\\_report.pdf](https://www.iftf.org/fileadmin/user_upload/images/Digital/IFTF_State_sponsored_trolling_report.pdf)
- [33] N. Ferenczi, T. C. Marshall, and K. Bejanyan, Are sex differences in antisocial and prosocial Facebook use explained by narcissism and relational self-construal?, *Computers in Human Behavior* 77, 2017, 25-31. <https://doi.org/10.1016/j.chb.2017.08.033>.
- [34] R. Regu, Changing forms and platforms of misogyny: Sexual harassment of women journalists on Twitter, *Media Watch* 9(3), 2018, 437-446.
- [35] F. Trompenaars, and C. Hampden-Turner, *Riding the waves of culture: Understanding diversity in global business*, 3rd edition. McGraw-Hill, 2012.
- [36] Y. B. R. Silintowe, and A. Susanto, Intercultural communication of Indian cultural dimensions through Indonesian secretary’s perspective, *Jurnal Ilmu Komunikasi*, 15 (2), 2012, 127-140.
- [37] P. Shachaf, and S. Horowitz, Are virtual reference services color blind? *Library & Information Science Research* 28(4), 2006, 501-520.