Title

Managing incivility in online brand communities

Denitsa Dineva and Jan Breitsohl

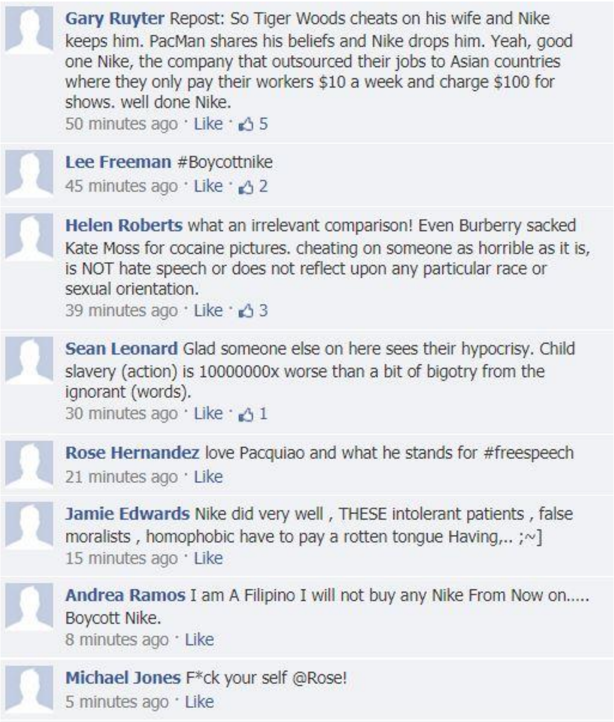
Abstract

Marketers have long utilized online brand communities to generate desirable user engagement behaviors. However, the rise of online brand communities has brought together millions of heterogenous users with diverse engagement motives, leading to the recent notion of 'the dark side' of social media engagement i.e., online incivility. This chapter offers an overview of this phenomenon. We first outline how incivility has been conceptualized in the literature, note a lack of terminological consensus, and propose some avenues for a possible theoretical integration. Drawing on findings across research disciplines, we then discuss different perspectives on how uncivil online interactions should be managed. Lastly, we combine these findings to offer a number of practical recommendations for digital and social media marketing managers and highlight three distinct avenues for future research.

Keywords: *online brand community; deviant online behaviors; incivility management; digital marketing management; interdisciplinary perspectives; social media ethics*

Introduction

Online brand communities are described as groups of consumers who express mutual sentiments about a particular brand, organization or consumption activity (Laroche, Habibi, Richard and Sankaranarayanan, 2012). The benefits of user interactions in these communities are well researched: individuals obtain social as well as functional value, while companies learn about consumer behaviors and market trends (e.g., Kim, Naylor, Sivadas and Sugumaran, 2016). However, there is a dark side to these communities. Online brand communities bring together millions of followers with heterogeneous socio-cultural backgrounds, belief systems and brand perceptions, and these differences increasingly lead to user-to-user interactions becoming hostile (i.e., online incivility) (Dineva, Breitsohl and Garrod, 2017). Unlike undesirable user-to-business interactions (e.g., complaints), these uncivil interactions neither typically originate in a product/service failure, nor do they demand a corporate remedy (Bacile, Wolter, Allen and Xu, 2018). Rather, online incivility represents interpersonal interactions between the followers of a brand community or the followers of opposing brand communities who use profanity, disagree with, provoke, harass or bully one another (Breitsohl, Roschk and Feyertag, 2018). Consequently, traditional forms of managing hostile interactions (e.g., monetary compensation, providing an apology) are unfit for this purpose and organizations have developed new strategies to address this increasingly prevalent phenomenon. To illustrate this, the excerpt in Figure 1 shows an uncivil interaction between the followers of Nike’s official Facebook community about the brand’s dismissal of celebrity endorser Manny Pacquiao following his derogatory comments about same-sex couples.



Gary Ruyter Repost: So Tiger Woods cheats on his wife and Nike keeps him. PacMan shares his beliefs and Nike drops him. Yeah, good one Nike, the company that outsourced their jobs to Asian countries where they only pay their workers $10 a week and charge $100 for shoes. well done Nike.
50 minutes ago · Like · 👍 5

Lee Freeman #Boycottnike
45 minutes ago · Like · 👍 2

Helen Roberts what an irrelevant comparison! Even Burberry sacked Kate Moss for cocaine pictures. cheating on someone as horrible as it is, is NOT hate speech or does not reflect upon any particular race or sexual orientation.
39 minutes ago · Like · 👍 3

Sean Leonard Glad someone else on here sees their hypocrisy. Child slavery (action) is 10000000x worse than a bit of bigotry from the ignorant (words).
30 minutes ago · Like · 👍 1

Rose Hernandez love Pacquiao and what he stands for #freespeech
21 minutes ago · Like

Jamie Edwards Nike did very well , THESE intolerant patients , false moralists , homophobic have to pay a rotten tongue Having,.. ;~]
15 minutes ago · Like

Andrea Ramos I am A Filipino I will not buy any Nike From Now on..... Boycott Nike.
8 minutes ago · Like

Michael Jones F*ck your self @Rose!
5 minutes ago · Like

Figure 1 Incivility excerpt

This particular interaction continued for four days on Nike’s community (over 35 million followers), generating 160 individual comments and 12,258 reactions. Research indicates that similar uncivil interactions in online brand communities can have a negative impact on the members of the community as well as adverse commercial outcomes for the brand in question (Bacile et al., 2018). More specifically, from a social perspective, uncivil online behaviors can increase hostility, decrease user well-being, cause mental distress and feelings of social isolation (Pew Research Center, 2017). Commercially, the outcomes of online incivility reduce self-brand identification and community engagement by both the victims and the observers of uncivil communications (Anderson et al., 2014; Moor, Heuvelman and Verleur, 2010). This further leads to brand reputational as well as financial loss (Adjei, Nowlin and Ang 2016; Ransbotham, Fichman, Gopal and Gupta, 2016).

Consequently, the management of incivility has attracted the attention of researchers from various disciplines and specifically mostly within management, marketing and IT among others. Based on the type of uncivil interaction and its severity, different coping mechanisms were put forward and these can be broadly divided into passive and active incivility management (Homburg, Ehm and Artz, 2015; Sibai, de Valck, Farrell and Rudd, 2015). While passive incivility management involves no intervention in the incident, active management consists of a range of verbal options that address the uncivil interaction (e.g., Dineva, Lu and Breitsohl, 2019; Matzat and Rooks, 2014). The focal point of this chapter represents an overview of contemporary incivility management practices that take place in online brand communities.

First, we review how prominent uncivil behaviors including flaming, trolling, cyber-bullying and consumer conflicts have been conceptualized in the literature. We outline possible reasons for why these occur, and the possible social as well as commercial consequences of these. We then move onto presenting recent research findings that highlight the defense mechanisms that digital marketers can use to tackle uncivil online behaviors. Finally, we offer a section on key take-aways for practitioners and illustrate several avenues for future research.

Uncivil online interactions and their impact

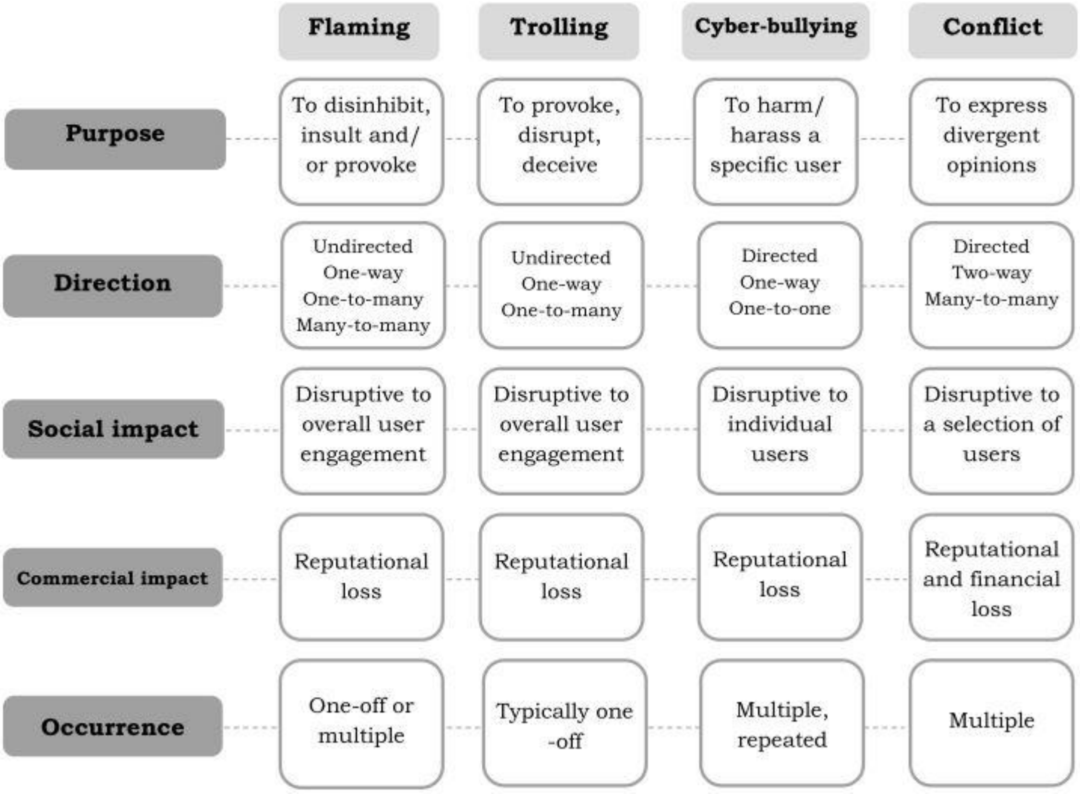
Online incivility refers to hostile and offensive communications between the members of online brand communities (Breitsohl et al., 2018; Dineva et al., 2017). The theoretical perspectives on the causes of online incivility suggest that a primary antecedent of uncivil online interactions is the relaxing of social and normative inhibitions and expectations typically inherent in face-to-face interactions (Suler, 2004, 2016). This is known as the disinhibition effect, which enables online users to freely communicate in an uncivil manner due to a temporary suspension of what is right or wrong in online settings.

Among several macro-level online characteristics, a main driving force of the disinhibition effect, recognized by Suler (2004) as well as others (Barlett and Gentile, 2012;

Lapidot-Lefler and Barak, 2012), is the full or partial anonymity when interacting with others online. Specifically, Li (2007) confirms that anonymity in computer-mediated communications leads to an increase in online deviance, because it not only reinforces disinhibition, but also reduces social accountability. The absence of social cues and social presence due to dissociative anonymity enables online community users to freely perform uncivil behaviors in the cyberspace that they feel inhibited from performing in offline environments (Lowry, Zhang, Wang and Siponen, 2016; Suler, 2004).

Deindividuation represents a secondary causal mechanism of online incivility fostered by anonymity and refers to losing one's sense of individuality and personal responsibility in the cyberspace (Valkenburg and Peter, 2011). According to Silke (2003), anonymity contributes to deindividuation through losing one's self-awareness from personal to the group. Consequently, online community members inhibit their personal sense of responsibility and convince themselves that they are not responsible for their communications or actions online (Freestone and Mitchell, 2004, Harris and Dumas, 2009) leading to occurrences of incivility. Additionally, in online communities, deindividuation amplifies the influence of group norms, which in turn can foster the learning and replication of online deviant behaviors (DeHue, Bolman and Völlink, 2008)

Prominent uncivil user-to-user interactions distinguished and examined by past research include: *flaming, trolling, cyber-bullying* and *consumer conflicts* (Breitsohl et al., 2018; Dineva et al., 2017; Herring, Job-Sluder, Scheckler and Barab, 2002; Lee, 2005; Lowry et al., 2016). Flaming, trolling and cyber-bullying are among the first uncivil online interactions to be captured by empirical research (Lowry et al., 2016; Herring et al., 2002; Lee, 2005). While traditionally these occurred on online forums, discussion boards, chatrooms and closed user-hosted communities, more recently they have transcended these boundaries and nowadays represent a commonplace on online brand communities. Moreover, since the rise of online brand communities another form of online incivility has emerged, which is centered around the brand and related consumption topics and issues – consumer conflicts (Breitsohl et al., 2018; Dineva et al., 2017). The main characteristics of these uncivil online interactions are summarized in Figure 2 and discussed in more detail in the following paragraphs.



	Flaming	Trolling	Cyber-bullying	Conflict
Purpose	To disinhibit, insult and/ or provoke	To provoke, disrupt, deceive	To harm/ harass a specific user	To express divergent opinions
Direction	Undirected One-way One-to-many Many-to-many	Undirected One-way One-to-many	Directed One-way One-to-one	Directed Two-way Many-to-many
Social impact	Disruptive to overall user engagement	Disruptive to overall user engagement	Disruptive to individual users	Disruptive to a selection of users
Commercial impact	Reputational loss	Reputational loss	Reputational loss	Reputational and financial loss
Occurrence	One-off or multiple	Typically one -off	Multiple, repeated	Multiple

Figure 2 Forms of incivility and their characteristics

Flaming represents a hostile expression of strong emotions such as swearing, insults, and name-calling, and according to Lee (2005), has been one of the most widely recognized forms of online incivility. In addition, flaming is a commonly employed linguistic tactic used to trigger emotional arousal and a sense of offensiveness (Jay and Janschewitz, 2008; Kwon and Cho, 2017). Thus, flaming can be categorized as intentional with the purpose to disinhibit, insult and/or provoke other community members. Flaming is considered by some as a distinct form of uncivil online communications (e.g., Anderson et al., 2014; Kwon and Gruzd, 2017), while others argue that it cannot be easily distinguished from other forms of incivility (Jay and

Janschewitz, 2008; Kenski, Coe and Rains, 2020). In relation to this, a seminal study characterized ‘true’ flaming as communications whereby the sender’s intent is to violate norms and both the receiver and any third-party observers perceive the message as a violation (O’Sullivan and Flanagin, 2003). The authors further proposed that flaming can be considered from the perspectives of the sender, receiver and third-party observer along a continuum ranging from mildly to highly inappropriate.

It has been confirmed that flaming causes significant disruption in online brand communities, because it spreads faster than non-offensive communications and thus reaches more users (Song et al., 2020). The consequences of this are two-fold: 1) with an increased number of community members being exposed to flaming, more users are likely to engage in hostile communications due to a ‘contagiousness effect’ (Kwon and Gruzd, 2017); and 2) some community members will disengage from the community and otherwise beneficial interactions with others (Chalmers Thomas, Price and Schau, 2013). Consequently, brands may experience a loss in credibility, if they fail to effectively deal with the flaming comments (Bacile et al., 2018).

Similarly to flaming, trolling represents a main form of online incivility, which is characterized as a deliberate behavior aimed at aggravating and disrupting others, but with no instrumental purpose (Buckels, Trapnell and Paulhus, 2014; Golf-Papez and Veer, 2017; Hardaker, 2010). The distinctive characteristic of this intentional and undirected at specific users behavior is the element of deception i.e., apparent outward sincerity by the sender of the message, the message is designed to attract flames, and waste the other users’ time by provoking futile arguments (Herring et al., 2002). Furthermore, based on a content analysis of online discussion forums, Hardaker (2010) identified four fundamental characteristics central to trolling behavior: deception (i.e., falsely portraying themselves), aggression (i.e., annoy and emotionally provoke others), disruption (i.e., meaningless distraction aimed at attentionseeking), and success (i.e., success in deceiving, aggravating, and disrupting the people they troll). In line with the latter, Craker and March (2016) found that negative social reward (i.e., having power and dominance over others) is among the strongest motivators of trolling behavior, followed by psychopathy and sadism from the Dark Tetrad of personality traits.

While Sanfilippo, Fichman and Yang (2018) differentiated between four trolling behavioral types ranging from humorous, non-serious trolling to non-humorous, serious trolling that is disruptive to online communities, the majority of empirical evidence points to the negative consequences of trolling (e.g., Binns, 2012; Thacker and Griffiths, 2012). Trolling disrupts otherwise constructive engagement on online brand communities and distracts users from engaging in meaningful interactions with like-minded supporters of the brand (Jiang et al., 2018; Phillips, 2011). Victims and bystanders often perceive this form of incivility as antisocial or deviant and report the same emotional and psychological outcomes as face-to-face forms of harassment such as depression, social anxiety, and low levels of self-esteem (Nicol, 2012).

In contrast to flaming and trolling, cyber-bullying involves repetition and power imbalance between the cyber-bully and the victim (Langos, 2012). While the three misbehaviors share similar attributes (i.e., aggression and intentionality) (Dooley, Pyżalski and Cross, 2010), cyber-bullying often involves a pre-existing relationship between the cyber-bully and the victim and this form of incivility is largely targeted (Steffgen, König, Pfetsch and Melzer, 2011). According to the literature, cyber-bullying consists of two other specialized forms i.e., cyber-stalking and cyber-harassment, but these are not always clearly distinguished (Lowry et al., 2016). Ultimately, however, studies on the different forms of cyber-bullying consistently demonstrate that individuals are more likely to engage in such behaviors online than offline (Slonje, Smith and Frisén 2013).

As cyberbullying in the social media can spread with a rapid, broad scale that it is almost unstoppable (Huang and Chou, 2010; Li, 2008), the consequences of cyber-bullying can be profound and affect all users of the community directly or indirectly and not just the victims (Pew Research Center, 2017). Direct encounters to cyber-bullying have offline consequences for the victims that range from mental and emotional distress to reputational damage and fear for one's personal safety. Indeed, scholars have recognized that due to the increased volume and scale as well as number of witnesses in online brand communities, cyber-bulling can cause greater emotional and psychological damage compared with traditional physical bullying (Gillespie, 2006). Indirectly, this form of online incivility causes social media users to refrain from positing online as well as some users discontinuing their use of social media after witnessing other community members being harassed or engaging in cyber-bullying behaviors (Camacho, Hassanein, and Head, 2018).

Lastly, consumer conflict has been captured by more recent studies and relates to one community member verbally attacking another who reciprocates the hostility (Breitsohl et al., 2018; Dineva et al., 2017). Thus, this form of online incivility entails a two-way exchange, unlike trolling and cyber-bullying, and is the outcome of different users disagreeing with and/or harassing each other specifically in relation to a brand or a consumption activity (unlike flaming). In addition, this type of incivility can revolve around disagreements related to others (i.e., the consequences of consumption on the environment) or related to the self (i.e., consuming to benefit the self solely) (Dineva, Breitsohl, Garrod and Megicks, 2020). Motivations behind engaging in this form of online deviance range from hostile to non-hostile ones (Breitsohl et al., 2018). While the former refers to strong language, high emotional intensity and adverse consequences for the brand such as loss of credibility, the latter involves humor and constructive criticism.

Importantly, research has confirmed that, if unmanaged, the accumulation of this type of incivility can generate online firestorms which can be detrimental to the brand's reputation and may result in financial losses (Hauser, Hautz, Hutter and Füller, 2017; Pfeffer, Zorbach and Carley, 2014). Moreover, others have demonstrated that individuals respond negatively to online incivility directed at them or their views (Phillips & Smith, 2004). King (2001) further showed that this type of incivility is strongly linked to affective responses by online users such as hatred and humiliation, and decreases perceptions of source and message credibility (Ng and Detenber, 2005). Additionally, when incivility targets an individual's ideological beliefs, it may influence the formation of negative attitudes about the issue/brand at hand (Hwang, Borah, Namkoong and Veenstra, 2008). For brands, this is detrimental since they are unable to effectively communicate promotional messages that facilitate desirable interactions between the members of their online brand communities. Ultimately, empirical research shows that online conflicts can negatively impact attitudes towards the consumption and adoption of products (Hansen, Kupfer and Hennig-Thurau, 2018).

In sum, while some scholars demonstrate that there may be a productive side to some forms of online incivility, the majority of research confirms that these largely produce negative outcomes for Internet users as well as brands operating online brand communities. While each of the discussed forms of incivility has its distinctive characteristics, they all share common attributes – malice, aggression and deliberation. It is important to also note that these forms of online incivility do not always operate in isolation or independent of one another. They are, in fact, not mutually exclusive and often one type of uncivil interaction is the antecedent or outcome of the other. For instance, trolling can often lead to a conflict, while conflicts frequently involve a degree of flaming.

Defense against the dark side

The literature on managing online incivility generally falls into the domain of *passive* versus *active* management approaches, as illustrated in Figure 3. Passive management involves community moderator behaviors such as avoiding the uncivil interactions (Hauser et al., 2017), remaining silent or ignoring the incivility (Godes et al., 2005; Hardaker, 2015), and observing without participating (Homburg et al., 2015). In contrast, active management consists of a range of verbalized community moderator practices that address the online incivility. As such, active management can be further grouped into *proactive* versus *reactive* approaches. A proactive approach relies on pre-defined community norms, formal rules, and expectations of community users to comply with these, while reactive management refers to addressing the incivility incident after it has occurred (Dineva et al., 2019). Reactive approaches can thus be divided into *positive* versus *negative* incivility management. Positive strategies involve a degree of cooperation and are aimed at encouraging desirable community behaviors and interactions, whereas negative management tends to be more assertive and is used to address more severe and harmful community behaviors (Matzat and Rooks, 2014). These perspectives on the management of online incivility have been offered by the management, marketing and IT literature, and here we focus on reviewing these as well as other notable studies from other disciplines.

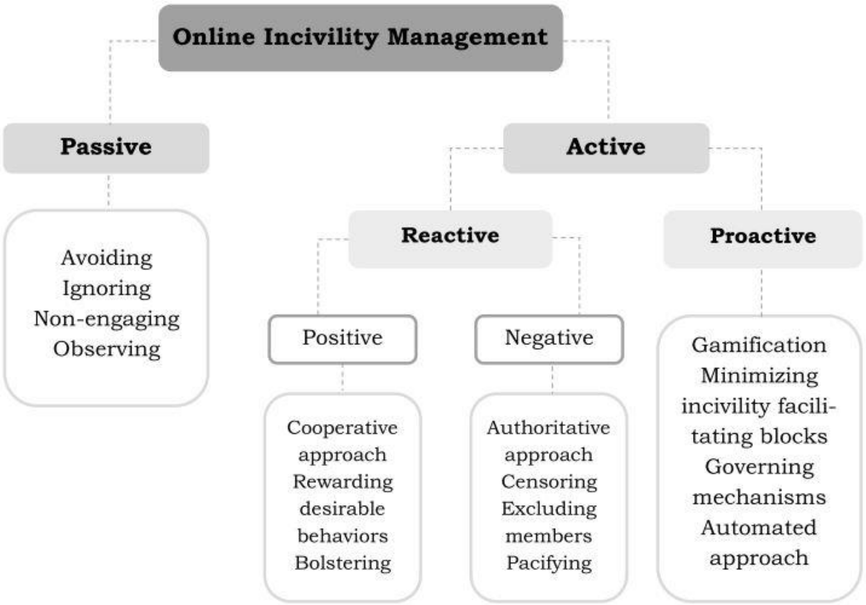


Figure 3 Online incivility management perspectives

Management perspectives

The management literature on dealing with online incivility is largely anecdotal. Most authors focus on giving practical accounts of incivility moderation, with the aim of identifying dos and don'ts (Fournier and Lee, 2009; Gallaugher and Ransbotham, 2010; Williams and Cothrel, 2000). Fournier and Lee (2009), for instance, use case studies from Dove, Apple and Porsche to illustrate that most companies choose to avoid engaging with incivility in their online brand communities. Williams and Cothrel (2000) suggest that successfully managing incivility in online communities requires explicit rules and formal moderation carried out by experienced moderators. In relation, Gallaugher and Ransbotham (2010) use Starbucks as a case study to provide general guidelines on managing uncivil online behaviors on firms' social media fan pages. The authors recommend that companies should opt for active content moderation, i.e., responding to uncivil online interactions without reinforcing negative behaviors. In addition, they put forward that companies should refrain from using censoring to moderate online incivility, because this is likely to exacerbate the issue. Likewise, in the context of online service recovery, Bacile (2020) emphasizes that importance of community managers to engage in active incivility management, because it is disruptive to how brands respond to customer complaints. This, in turn, the study's findings suggest generates negative customer perceptions of the digital service environment and harm the customer experience.

Two seminal models of conflict management from the organizational behavior literature are widely discussed in more recent studies on managing incivility in online environments – Blake and Mouton's (1964) managerial grid and Rahim's (1983) conflict management model. The models offer a consistent pattern of conflict management styles (i.e., integrating, obliging, sharing, dominating and avoiding) that fall into a broader set of categories – cooperative, assertive and avoiding conflict management. While cooperative strategies refer to an open-minded discussion with a focus on understanding the opposing arguments, assertive management styles involve defending one's position and pursuing one's own interest at the expense of others (Rahim, 2002). An avoiding management approach, in contrast, is associated with no intervention in the uncivil interactions. Generally, past research has concluded that a cooperative approach is preferred to assertive or avoiding styles of incivility management (Antoci et al., 2016; Ishii, 2010).

Focusing on managing the overall online community environment, rather than utilizing specific strategies to combat online incivility, O'Mahony (2007) identified five mechanisms that can minimize the negative impact of online incivility when it occurs. These include: independence (issues are resolved between community members; a formal governance body is not required), pluralism (multiple points of view are considered), representation (incivility management is democratically assigned to some community members), decentralized decision-making (formal governance body is present, but some problem-solving rights are distributed to community members) and autonomous participation (freedom is given to community members to contribute on their own terms). In a similar vein, based on a review of the literature, Fombelle et al. (2020) put forward generic mechanisms that brands can utilize in order to prevent online incivility from occurring. The authors propose that a combination of social as well as technological means are necessary to combat deviant online behaviors such as trolling and anti-brand offensive communications. While social mechanisms represent establishing and reinforcing social norms of expected online behaviors and compliance with these (Schaefers, Wittkowski, Benoit and Ferraro, 2016), technological means include software algorithms and chatbots that interact, predict and flag trolls and online aggression as well as ban suspicious users (Reynolds, Kontostathis and Edwards, 2011).

Marketing perspectives

Godes et al. (2005) offered first insights into roles that companies may adopt when managing user-to-user interactions. The authors distinguished between four principal, non-mutually exclusive company roles ranging from passive observation to interactive participation. Depending on the type and severity of the interaction and the context, the company can choose between the following roles: participant, moderator, mediator and observer. Likewise, in interacting with the users of an online community and managing user-to-user discussions, Homburg et al. (2015) identified passive and active company roles. Passive engagement entails the company offering individuals a platform to interact and refrains from engaging in conversations among them, while active participation involves conversing with the participants and intervening in uncivil discussions.

In the field of consumer research, a study by Sibai et al. (2015) argued that the heterogeneity of online consumption communities requires managers to exercise social control through governance structures and moderation practices. The authors put forward two strategies to manage conflicts. On the one hand, interaction maintenance follows a proactive, ongoing approach, which involves explicating roles, formalizing rules, monitoring interactions, rewarding positive behaviors and sanctioning negative behaviors. For instance, explicating roles pre-defines positions or functions that have corresponding responsibilities in managing the conflict. Similarly, formalizing rules specifies rights to be used for future contingencies. In contrast to these, monitoring refers to keeping records of behavior in order to understand the causes of the incivility. Rewarding or sanctioning behavior represents a set of actions that incentivize positive behavior or dismiss incentives for negative behavior. On the other hand, interaction termination is more reactive in nature and seeks to end interactions that have become dysfunctional either by ignoring members or by permanently excluding them from the community. In line with this, Husemann, Ladstaetter and Luedicke (2015) propose that community moderators should adopt assertiveness in managing incivility, for example, through excluding members from the community. This proposition is based on the authors distinguishing between routinized (i.e., constructive) and transgressive (i.e., uncivil) interactions in online brand communities and acknowledging that the latter are destructive to the well-being and engagement of community members.

Two previous studies on conflict management within user-hosted online brand communities have put forward the concept of community-governing mechanisms (Mathwick, Wiertz and De Ruyter, 2007; Schau, Muñiz and Arnould, 2009). These constitute articulating expectations for acceptable behavior including maintaining criticism constructive, dismissing unjustified, negative comments, and sustaining a positive community environment. Specifically, Schau et al. (2009) recommend articulating expectations for acceptable behavior, followed by dismissing ‘flaming’ comments and/or unjustified criticism in the community as the most common governing mechanism in online communities. Similarly, in the context of trolling, Golf-Papez and Veer (2017) argue that the effective combatting of incivility lies with managing its building blocks and facilitating factors as opposed to addressing individual acts of incivility. Consequently, the authors propose that minimizing or eliminating the effects associated with trolling including building awareness, reducing provocations, rewarding the reporting of trolling and enforcing sanctions are effective mechanisms for dealing with online incivility.

In contrast, others have outlined typologies of managing incivility on online brand communities explicitly based on the management of individual occurrences of incivility (Dineva et al., 2017; Dineva et al., 2020). In commercially oriented online brand communities, Dineva et al. (2017) distinguished between verbal and non-verbal conflict management strategies. The former consist of bolstering (i.e., reinforcing the comments made by brand defenders), pacifying (i.e., asking participants in an uncivil interaction to adjust their communications), and informing (i.e., correcting misinformation), while the latter involve not engaging in the incivility incidents and censoring uncivil comments. A later study uncovered similar strategies utilized in non-profit online brand communities as well as an additional strategy specific to the non-profit consumption context – mobilizing i.e., urging the users involved in the uncivil interactions to change their stance or consumption behaviors (Dineva et al., 2020). In addition, the study's findings confirmed that the pacifying, mobilizing and bolstering strategies generate the most favorable user attitudes towards the organization/brand in question, while censoring and non-engaging negatively impact user perceptions of the organization's social responsibility efforts.

IT management perspectives

In the information technology (IT) literature, Matzat and Rooks (2014) drew a contrast between positive and negative approaches to managing online incivility. Similarly to other studies, the authors conceptualized positive conflict management as rewarding desirable behaviors, while negative conflict management is described as punishing uncivil behaviors and interactions. In relation, Huang et al.'s study (2016) exclusively focused on positive approaches to managing incivility and examined the effectiveness of three strategies – rational explanation, constructive suggestion, and social encouragement on individuals' willingness to remain in and contribute to a collaborative consumption-based online community. Rational explanation is described as a reactive, focused on the issue approach that involves providing more detailed information and clarifying misunderstandings among the conflicting parties. Social encouragement, in contrast, is deemed as more proactive and aims to create a friendly online environment to prevent incivility from occurring. Finally, constructive suggestion is most commonly used and refers to suggesting concrete alternative solutions to the incivility incident. The study's findings further point to constructive suggestion as most effective when resolving incivility since it facilitates community members retention.

In information management, borrowed from the organizational behavior literature (Rahim, 1983), Hauser et al. (2017) investigated the effect of assertive versus cooperative approaches on addressing the accumulation of conflict and aggression in online brand communities. Assertive conflict management is represented by competing, obliging and avoiding, which were found to further escalate the conflict. On the contrary, cooperative conflict management involves accommodating, yielding, integrating strategies, which can be described as showing willingness to cooperate with the opposing party. The study concludes that cooperative strategies are generally more effective, but acknowledges that the success of assertive versus cooperative strategies is dependent upon other factors such as attitudes towards the community, the number and presence of moderators and their perceived credibility.

In a similar vein, specifically in response to trolling behaviors, Sanfilippo et al. (2017) suggested that there are certain factors that need to be considered in determining an appropriate trolling management strategy including the context, the platform and whether the act of trolling itself is perceived as deviant. Based on these, companies can choose between simply ignoring the troll, or implementing stricter measures such as blocking the trolls and/or deleting their posts as well as unmasking their identities. In line with considering the context and factors surrounding online incivility, Cruz, Seo and Rex (2018) confirmed that there are certain social practices within online communities that enable the formation of trolling behaviors (i.e., learning, assimilation and transgression). Consequently, the authors suggested that managers of online communities should focus on proactively managing the social practices that enable this online misbehavior, rather than addressing the individual trolling instances.

Perspectives from other disciplines

Several more studies found in the politics, sociology, cyber-psychology, journalism and communications literature are also relevant and discussed here. In politics, Wright's (2006) work on online forums run by the government differentiates between content moderation and interactive moderation in a similar fashion to Sibai et al.'s (2015) study on consumption communities. Content moderation is characterized by content removal and the absence of justification for the deletion. Interactive moderation, in contrast, represents two-way communication between the moderator(s) and the community members and includes maintaining civility and encouraging thorough discussions. In line with this positive versus negative incivility moderation paradigm, in journalism Binns (2012) put forward 'gamification'. The approach involves adopting video games techniques in non-gaming contexts such as online brand communities, for example, to address trolling. Thus, through 'gamification' a community moderator rewards and encourages desirable community behaviors, for example, through awarding tokens and reducing the anonymity of the users in the community. This, in turn, is expected to facilitate an overall positive online environment and discourage uncivil behaviors from taking place. A related communications study on managing trolling in online newsgroups recommended several techniques (Hardaker, 2015). These techniques adopted by the users of the online community include exposing the troll to other users, challenging the troll, criticizing the effectiveness/success of the trolling, mocking or parodying the trolling attempt, and reciprocating in kind by trolling the troll.

In sociology, Lee (2005) outlined behavioral strategies used in a feminist online forum to deal with flaming among its members. Following early work by Oetzel et al. (2000), Lee's (2005) strategies are categorized into three groups: competitive-dominating, cooperative-integrating and avoiding. The competitive-dominating strategies involve threats, persuasion and requesting compliance, whereas the cooperative-integrating approach suggests an overall consideration of others, including compromising, offering concessions, apologizing and showing solidarity. Avoiding strategies, in contrast, comprise of activities that aim to ignore the conflict, including making jokes, being silent, bringing in third parties and withdrawal. In another early sociological study, Smith (2002) offered three main mechanisms for social control when online incivility occurs – mediation, fact-finding and arbitration. While mediation refers to neutral negotiation that facilitates an agreement between the disputants, fact-finding and arbitration appear to be more authoritative. Specifically, fact-finding relies on resolving the conflict through determining the facts and rejecting the meritless argument and arbitration provides the final resolution without necessarily considering opposing views. Smith (2002) further adds that arbitration is frequently the least preferred option, while mediation and fact-finding are more effective in preventing incivility from escalating and sustaining constructive interactions between community users.

A study in cyber-psychology by Mishna et al. (2011) conducted a review of studies on online intervention and prevention programs to address cyber-bullying in the social media. Their findings confirmed that an automated approach which involves developing and using technological and software initiatives that block and filter incivility and misconducts is appropriate for effectively addressing cyber-bullying. Although reliance on automated detection and management of online incivility may be suited in some instances, it may not always be the only appropriate option. Anderson, Bresnahan and Musatics (2014), for instance, introduced a model of dissenting behavior to serve as a cyber-bullying prevention tool in addition to the automated approach. Their study suggested that disagreeing with the aggressor will encourage more bystanders to provide social support to the victim.

Practical implications

Recent research consistently demonstrates that when incivility occurs in online brand communities, community hosts predominantly do not intervene (Bacile et al., 2018; Breitsohl et al., 2018). This passive approach to managing uncivil online interactions may be a preferred option due to it being unobtrusive, cost-effective and resource non-intensive (Dineva et al., 2017). Indeed, it may be easier for online community managers to simply dismiss incivility as unharmful and humorous teasing behaviors, and passively monitoring uncivil interactions may be suited in some instances. Most empirical research demonstrates that due to its nature and characteristics, ignoring trolling behaviors, for example, can be an effective approach, as it deprives the troll(s) from attention (e.g., Hardaker, 2015). Nonetheless, community moderators should first determine that the uncivil interaction is harmless and merely humor oriented before opting for non-engagement (Breitsohl et al., 2018; Sanfilippo et al., 2018). In the majority of incivility instances, however, lack of involvement is not a viable approach (Dineva et al., 2020). This is because users of online brand communities hold certain expectations of community management and attribute responsibility to the company/brand for ensuring civil and moderated engagement when incivility takes place (Henkel, Boegershausen, Rafaeli and Lemmink, 2017).

User engagement in online brand communities takes place in a fast-paced, global environment and is facilitated by automated algorithms. These algorithms are designed to promote comments, threads and posts that receive high engagement rates from other users, which inevitably attracts and accelerates incivility (Ilhan, Kübler and Pauwels, 2018). For this reason, it becomes an imperative for online brand community moderators to establish and act on pre-defined community rules for civil engagement. Research to date has demonstrated that such a proactive approach helps reinforce desirable behaviors and creates a positive engagement environment that naturally minimizes incivility (Binns, 2012; Schau et al., 2009). Proactively informing community members of desired behaviors, while highlighting the community's lack of tolerance for harmful ones enables community managers to assert their authority with no repercussions for the brand when community rules are breached, and comments are censored, or members of the community are removed. Facebook, for instance, allows business pages to proactively moderate user content through blocking the use of certain words and using profanity filters (Facebook for Business, 2021). Twitter, in contrast, relies on various inappropriate content policies that guide business accounts (Twitter for Business, 2021), but moderation is predominantly reliant on users reporting inappropriate content or 'muting' keywords, tweets and accounts.

Proactive incivility management alone however is not sufficient in effectively dealing with uncivil online interactions. Digital marketers thus have a choice of reactive strategies when dealing with individual instances of incivility. These are broadly categorized into positive versus negative incivility management. A positive approach includes strategies such as positively reinforcing a brand defender, rewarding desirable community behaviors and providing further information to the involved parties among others (Dineva et al., 2017). In contrast, a negative approach requests compliance from the users who engage in online incivility through asking them to adjust their communication behavior or style and censoring their comments or removing them from the community in more severe instances (Bacile et al., 2018). When choosing suitable strategies for their online brand communities, it is important to note that positively oriented strategies are generally received more favorably by community members, while more assertive strategies are perceived as less effective (Hauser et al., 2017). Facebook offers brands another means to reactively moderate incivility. Among several non-verbal approaches are temporarily hiding comments, deleting comments, and temporarily or permanently suspending a user from the brand community (Facebook for Business, 2021).

Where do we go from here?

The existing literature suggests that community managers should employ a set of strategies to exercise social control when uncivil online interactions take place. Based on these insights from various conceptual and empirical perspectives, we propose three main avenues for future research.

Research avenue 1: Communication content

Future studies should focus on the content of organizational communication strategies. While recent empirical research indicates that many companies currently tend to remain inactive during uncivil online interactions, it is suggested that systematic observations of a broad range of online brand communities are necessary to expand upon the examples of current practice illustrated in past studies. Once a more generalizable overview of current practice across industries as well as cultures can be drawn, subsequent experimental research should further verify their effectiveness. While there is a general preference towards the positive/cooperative approaches to incivility management over more assertive strategies, studies should further test and confirm whether this notion is valid across different cultures and industries, and whether additional message framing manipulations might have a positive effect. For instance, research on message congruity in the e-complaint management literature shows that interventions that match the tone of one or several parties tend to yield more positive outcomes (Breitsohl, Khammash & Griffiths, 2010).

Likewise, theories of persuasion such as the elaboration likelihood model (Petty and Cacioppo, 1986) may be used to frame an intervention message based on the ability and motivation of the uncivil users. For instance, communication style (i.e., formal versus informal) has received significant attention by online communications researchers due to its ability to influence user perceptions (Javornik, Filieri and Gumann, 2020; Schamari and Schaefers, 2015). Future studies should investigate whether conversational/informal communication style in managing incivility generates more positive user responses (e.g., minimizes aggression) and perceptions of the brand (e.g., improved social responsibility) compared with a more corporate/formal style of addressing uncivil interactions.

Research avenue 2: Communication impact

To understand the effectiveness of manipulating the content of incivility management strategies, future research needs to further investigate commercial, social and policy impact factors. Studies have begun to examine these consequences, but these are currently at an infant stage with findings requiring further validation and generalization. Commercially, online brand communities will benefit from research that verifies which strategies have the most positive effect on consumers' brand relationship, organizational image, trust perceptions and loyalty-related behaviors. For instance, Chalmers Thomas et al. (2013) suggest that not intervening in uncivil interactions can escalate the severity of these and result in members leaving the online community. Of similar interest for future research is the effect of different strategies on the social well-being of online community members. To this regard, intervening in hostile interactions may enhance consumers' trust in social discourse online and prevent the negative emotional contagion of online incivility (Breitsohl et al. 2018).

Since online incivility has a profound social and commercial impact, its management has been raised by some scholars as a key area for consideration by policy makers (Brown, Jackson and Cassidy, 2006). While some suggest that the responsibility for managing online incivility lies with the companies/brands that host online communities (Dholakia, Blazevic, Wiertz and Algesheimer, 2009), the popular press seems to suggest that social media platforms should also be held accountable for addressing and minimizing its negative consequences on the society (e.g., BBC 2020; The Guardian, 2020). Future research should thus carry out discourse analysis into the processes involved in policymaking regarding the governance of social media platforms and online brand communities.

Research avenue 3: Communication context

Closely linked to content and impact, considerable research opportunities lay ahead in exploring boundary conditions, which reflect differences in the communication context. First, based on the review of the incivility management literature, an investigation into whether the effectiveness of the strategies depends upon the type of uncivil interaction is necessary. Since different forms of online incivility vary in their degree of aggression (Breistohl et al. 2018) and may at times actually prove constructive or unharmful (Husemann et al. 2015), the type of uncivil interaction may be an important moderator of the impact of management interventions. Second, the effectiveness of a strategy will vary in relation to the sender and the receiver of an online intervention. Based on social agency theories (Hartmann et al., 2008), it is likely that online community users will react differently depending on whether an intervention is posted by a brand, employee or brand advocate. Future studies may investigate, for instance, the combined effect of a community moderator and a brand advocate on user perceptions of civility and impact on continuous engagement in the community. Similarly, the effectiveness of an intervention may be different for an uninvolved witness or a bystander and an active participant in the uncivil interaction. Gretry, Horváth, Belei and van Riel (2017), for instance, showed that consumers perceive the organization as competent, if assertiveness is directed at others who engage in incivility, but not at them.

Conclusion

Overall, knowledge on how social media and digital marketing managers may handle incivility in their online brand communities is still in its infancy. While a good number of netnographic observations show that uncivil interactions present an increasingly imminent managerial challenge, the literature still lacks an accepted conceptualization of what incivility actually encompasses, and to what extent it causes harm. Recent studies have started to offer experimental insights on the commercial consequences of inter-consumer incivility, yet many gaps remain to be explored. This chapter aims to offer a first step towards an integrated theoretical foundation, aimed at encouraging future research to deepen our currently fragmented understanding of a key research topic in digital and social media marketing.

References

Adjei, M. T., Nowlin, E. L., & Ang, T. (2016). The collateral damage of C2C communications on social networking sites: the moderating role of firm responsiveness and perceived fairness. *Journal of Marketing Theory and Practice*, *24*(2), 166-185.

Anderson, J., Bresnahan, M., & Musatics, C. (2014). Combating weight-based cyberbullying on Facebook with the dissenter effect. *Cyberpsychology, Behavior, and Social Networking*, *17*(5), 281-286.

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373-387.

Antoci, A., Delfino, A., Paglieri, F., Panebianco, F., & Sabatini, F. (2016). Civility vs. incivility in online social interactions: An evolutionary approach. *PloS one*, *11*(11), e0164286.

Bacile, T. J. (2020). Digital customer service and customer-to-customer interactions: investigating the effect of online incivility on customer perceived service climate. *Journal of Service Management*, *30*(3), 441-464.

Bacile, T. J., Wolter, J. S., Allen, A. M., & Xu, P. (2018). The effects of online incivility and consumer-to-consumer interactional justice on complainants, observers, and service providers during social media service recovery. *Journal of Interactive Marketing*, *44*, 60-81.

Barlett, C. P., & Gentile, D. A. (2012). Attacking others online: The formation of cyberbullying in late adolescence. *Psychology of Popular Media Culture*, *1*(2), 123.

BBC (2020). Social media: How might it be regulated?, (Accessed: 7 December 2020), Available at: https://www.bbc.co.uk/news/technology-54901083

Binns, A. (2012). DON'T FEED THE TROLLS! Managing troublemakers in magazines' online communities. *Journalism Practice*, *6*(4), 547-562.

Blake, R. R., & Mouton, J. S. (1964). *The managerial grid. Houston*, TX: Gulf.

Breitsohl, J., Khammash, M., & Griffiths, G. (2010). E-business complaint management: perceptions and perspectives of online credibility. *Journal of Enterprise Information Management*, *23*(5), 653-660.

Breitsohl, J., Roschk, H., & Feyertag, C. (2018). Consumer brand bullying behaviour in online communities of service firms. In M. Bruhn & H. Karsten (Eds.) *Service Business Development* (pp. 289 – 289). Springer Gabler: Wiesbaden.

Brown, K., Jackson, M., & Cassidy, W. (2006). Cyber-bullying: Developing policy to direct responses that are equitable and effective in addressing this special form of bullying. *Canadian Journal of Educational Administration and Policy*, (57).

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97-102.

Camacho, S., Hassanein, K., & Head, M. (2018). Cyberbullying impacts on victims' satisfaction with information and communication technologies: The role of perceived cyberbullying severity. *Information & Management*, *55*(4), 494-507.

Chalmers Thomas, T., Price, L. L., & Schau, H. J. (2013). When differences unite: Resource dependence in heterogeneous consumption communities. *Journal of Consumer Research*, *39*(5), 1010-1033.

Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, *102,* 79-84.

Cruz, A. G. B., Seo, Y., & Rex, M. (2018). Trolling in online communities: A practice-based theoretical perspective. *The Information Society*, *34*(1), 15-26.

DeHue, F., Bolman, C., & Völlink, T. (2008). Cyberbullying: Youngsters' experiences and parental perception. *CyberPsychology & Behavior*, *11*(2), 217-223.

Dholakia, U. M., Blazevic, V., Wiertz, C., & Algesheimer, R. (2009). Communal service delivery: How customers benefit from participation in firm-hosted virtual P3 communities. *Journal of Service Research*, *12*(2), 208-226.

Dineva, D., Breitsohl, J. C., & Garrod, B. (2017). Corporate conflict management on social media brand fan pages. *Journal of Marketing Management*, *33*(9-10), 679-698.

Dineva, D., Breitsohl, J., Garrod, B., & Megicks, P. (2020). Consumer responses to conflict-management strategies on non-profit social media fan pages. *Journal of Interactive Marketing*, *52*, 118-136.

Dineva, D., Lu, X., & Breitsohl, J. (2019). Social media conflicts during the financial crisis: Managerial implications for retail banks. *Strategic Change*, *28*(5), 381-386.

Dooley, J. J., Pyżalski, J., & Cross, D. (2009). Cyberbullying versus face-to-face bullying: A theoretical and conceptual review. *Zeitschrift für Psychologie/Journal of Psychology*, *217*(4), 182-188.

Facebook for Business (2021). Admin’s Guide to Moderating Your Page, Retrieved from: https://en-gb.facebook.com/business/a/page-moderation-tips

Fombelle, P. W., Voorhees, C. M., Jenkins, M. R., Sidaoui, K., Benoit, S., Gruber, T., Gustafsson, A., & Abosag, I. (2020). Customer deviance: A framework, prevention strategies, and opportunities for future research. *Journal of Business Research*, *116*, 387-400.

Fournier, S., & Lee, L. (2009). Getting brand communities right. *Harvard Business Review, 87*(4), 105–111

Freestone, O., & Mitchell, V. (2004). Generation Y attitudes towards e-ethics and internet-related misbehaviours. *Journal of Business Ethics*, *54*(2), 121-128.

Gallaugher, J., & Ransbotham, S. (2010). Social media and customer dialog management at Starbucks. *MIS Quarterly Executive*, *9*(4), 197-212.

Gillespie, A. A. (2006). Cyber-bullying and harassment of teenagers: The legal response. *Journal of Social Welfare & Family Law, 28*(2), 123-136.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, Libai., L., Sen, S., Shi, M. & Verlegh, P. (2005). The firm's management of social interactions. *Marketing Letters*, *16*(3-4), 415-428.

Golf-Papez, M., & Veer, E. (2017). Don’t feed the trolling: rethinking how online trolling is being defined and combated. *Journal of Marketing Management*, *33*(15-16), 1336-1354.

Gretry, A., Horváth, C., Belei, N., & van Riel, A. C. (2017). “Don't pretend to be my friend!” When an informal brand communication style backfires on social media. *Journal of Business Research*, *74*, 77-89.

Hansen, N., Kupfer, A. K., & Hennig-Thurau, T. (2018). Brand crises in the digital age: The short-and long-term effects of social media firestorms on consumers and brands. *International Journal of Research in Marketing, 35*(4), 557-574.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research, 6*(2), 215-242.

Hardaker, C., (2015). ‘I refuse to respond to this obvious troll’: an overview of responses to (perceived) trolling”, *Corpora*, *10*(2), 201-229.

Harris, L. C., & Dumas, A. (2009). Online consumer misbehaviour: an application of neutralization theory. *Marketing Theory*, *9*(4), 379-402.

Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K., & Tucker, C. (2008). Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Letters*, *19*(3-4), 287-304.

Hauser, F., Hautz, J., Hutter, K., & Füller, J. (2017). Firestorms: Modeling conflict diffusion and management strategies in online communities. *The Journal of Strategic Information Systems*, *26*(4), 285-321.

Henkel, A. P., Boegershausen, J., Rafaeli, A., & Lemmink, J. (2017). The social dimension of service interactions: Observer reactions to customer incivility. *Journal of Service Research*, *20*(2), 120-134.

Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing" trolling" in a feminist forum. *The Information Society*, *18*(5), 371-384.

Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research*, *52*(5), 629-641.

Huang, Y. Y., & Chou, C. (2010). An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Computers in Human Behavior*, *26*(6), 1581-1590.

Huang, W., Lu, T., Zhu, H., Li, G., & Gu, N. (2016). Effectiveness of conflict management strategies in peer review process of online collaboration projects. In *Proceedings of the*

19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 717-728). ACM.

Husemann, K. C., Ladstaetter, F., & Luedicke, M. K. (2015). Conflict culture and conflict management in consumption communities. *Psychology & Marketing*, *32*(3), 265-284.

Hwang, H., Borah, P., Namkoong, K., & Veenstra, A. (2008, May). Does civility matter in the blogosphere? Examining the interaction effects of incivility and disagreement on citizen attitudes. In *58th Annual Conference of the International Communication Association, Montreal, QC, Canada*.

Ilhan, B. E., Kübler, R. V., & Pauwels, K. H. (2018). Battle of the brand fans: impact of brand attack and defense on social media. *Journal of Interactive Marketing*, *43*, 33-51.

Ishii, K. (2010). Conflict management in online relationships. *Cyberpsychology, Behavior, and Social Networking*, *13*(4), 365-370.

Javornik, A., Filieri, R., & Gumann, R. (2020). “Don't Forget that Others Are Watching, Too!” The Effect of Conversational Human Voice and Reply Length on Observers' Perceptions of Complaint Handling in Social Media. *Journal of Interactive Marketing*, *50*, 100-119.

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research*, *4*(2), 267-288.

Jiang, L., Mirkovski, K., Wall, J. D., Wagner, C., & Lowry, P. B. (2018). Proposing the core contributor withdrawal theory (CCWT) to understand core contributor withdrawal from online peer-production communities. *Internet Research*, 28(4), 988-1028.

Kenski, K., Coe, K., & Rains, S. A. (2020). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research*, *47*(6), 795-814.

Kim, J., Naylor, G., Sivadas, E., & Sugumaran, V. (2016). The unrealized value of incentivized eWOM recommendations. *Marketing Letters*, *27*(3), 411-421.

King, A. B. (2001). Affective dimensions of Internet culture. *Social Science Computer Review*, *19*(4), 414-430.

Kwon, K. H., & Cho, D. (2017). Swearing effects on citizen-to-citizen commenting online: A large-scale exploration of political versus nonpolitical online news sites. *Social Science Computer Review*, *35*(1), 84-102.

Kwon, K. H., & Gruzd, A. (2017). Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to Donald Trump’s YouTube campaign videos. *Internet Research*, 27(4), 991-1010.

Langos, C. (2012). Cyberbullying: The challenge to define. *Cyberpsychology, Behavior, and Social Networking*, *15*(6), 285-289.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, *28*(2), 434-443.

Laroche, M., Habibi, M. R., Richard, M. O., & Sankaranarayanan, R. (2012). The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Computers in Human Behavior*, *28*(5), 1755-1767.

Lee, H. (2005). Behavioral strategies for dealing with flaming in an online forum. *The Sociological Quarterly*, *46*(2), 385-403.

Li, Q. (2007). New bottle but old wine: A research of cyberbullying in schools. *Computers in Human Behavior*, *23*(4), 1777-1791.

Li, Q. (2008). A cross-cultural comparison of adolescents' experience related to cyberbullying. *Educational Research*, *50*(3), 223-234.

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, *27*(4), 962-986.

Mathwick, C., Wiertz, C., & De Ruyter, K. (2008). Social capital production in a virtual P3 community. *Journal of Consumer Research*, *34*(6), 832-849.

Matzat, U., & Rooks, G. (2014). Styles of moderation in online health and support communities: An experimental comparison of their acceptance and effectiveness. *Computers in Human Behavior, 36*, 65-75.

Mishna, F., Cook, C., Saini, M., Wu, M. J., & MacFadden, R. (2011). Interventions to prevent and reduce cyber abuse of youth: A systematic review. *Research on Social Work Practice, 21*(1), 1-10.

Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on youtube. *Computers in Human Behavior, 26*(6), 1536-1546.

Nicol, S. (2012). Cyber-bullying and trolling. *Youth Studies Australia*, *31*(4), 3-4.

Ng, E. W., & Detenber, B. H. (2005). The impact of synchronicity and civility in online political discussions on perceptions and intentions to participate. *Journal of Computer-Mediated Communication*, *10*(3), JCMC1033.

Oetzel, J. G., Ting-Toomey, S., Yokochi, Y., Masumoto, T., & Takai, J. (2000). A typology of facework behaviors in conflicts with best friends and relative strangers. *Communication Quarterly, 48*(4), 397-419.

O'Mahony, S. (2007). The governance of open source initiatives: What does it mean to be community managed? *Journal of Management and Governance, 11*(2), 139-150.

O'Sullivan, P. B., & Flanagin, A. J. (2003). Reconceptualizing 'flaming'and other problematic messages. *New Media & Society*, *5*(1), 69-94.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and Persuasion* (pp. 1-24). Springer, New York, NY.

Pew Research Center (2017). Online Harassment. (Accessed: 7 December, 2020), Available at: http://www.pewinternet.org/2017/07/11/online-harassment-2017/

Pfeffer, J., Zorbach, T., & Carley, K. M. (2014). Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. *Journal of Marketing Communications*, *20*(1-2), 117-128.

Ransbotham, S., Fichman, R. G., Gopal, R., & Gupta, A. (2016). Special section introduction—ubiquitous IT and digital vulnerabilities. *Information Systems Research*, *27*(4), 834-847.

Phillips, W. (2011). LOLing at tragedy: Facebook trolls, memorial pages and resistance to grief online. *First Monday*, 16(2).

Phillips, T., & Smith, P. (2004). Emotional and behavioural responses to everyday incivility: Challenging the fear/avoidance paradigm. *Journal of Sociology*, *40*(4), 378-399.

Rahim, M. A. (1983). A measure of styles of handling interpersonal conflict. *Academy of Management Journal*, *26*(2), 368-376.

Rahim, M.A. (2002). Toward a theory of managing organisational conflict. *The International Journal of Conflict Management*, *13*(3), 206-235.

Reynolds, K., Kontostathis, A., & Edwards, L. (2011, December). Using machine learning to detect cyberbullying. In *2011 10th International Conference on Machine learning and applications and workshops* (Vol. 2, pp. 241-244). IEEE.

Sanfilippo, M. R., Fichman, P., & Yang, S. (2018). Multidimensionality of online trolling behaviors. *The Information Society*, *34*(1), 27-39.

Schaefers, T., Wittkowski, K., Benoit, S., & Ferraro, R. (2016). Contagious effects of customer misbehavior in access-based services. *Journal of Service Research*, *19*(1), 3-21.

Schamari, J., & Schaefers, T. (2015). Leaving the home turf: How brands can use webcare on consumer-generated platforms to increase positive consumer engagement. *Journal of Interactive Marketing*, *30*, 20-33.

Schau, H. J., Muñiz Jr, A. M., & Arnould, E. J. (2009). How brand community practices create value. *Journal of Marketing*, *73*(5), 30-51.

Sibai, O., De Valck, K., Farrell, A. M., & Rudd, J. M. (2015). Social control in online communities of consumption: A framework for community management. *Psychology & Marketing*, *32*(3), 250-264.

Silke, A. (2003). Deindividuation, anonymity, and violence: Findings from Northern Ireland. *The Journal of Social Psychology*, *143*(4), 493-499.

Slonje, R., Smith, P. K., & Frisén, A. (2013). The nature of cyberbullying, and strategies for prevention. *Computers in Human Behavior*, *29*(1), 26-32.

Smith, A. D. (2002). Problems of conflict management in virtual communities. In *Communities in Cyberspace* (pp. 145-174). London: Routledge.

Song, Y., Kwon, K. H., Xu, J., Huang, X., & Li, S. (2020). Curbing profanity online: A network-based diffusion analysis of profane speech on Chinese social media. *New Media & Society*, 1-22.

Steffgen, G., König, A., Pfetsch, J., & Melzer, A. (2011). Are cyberbullies less empathic? Adolescents' cyberbullying behavior and empathic responsiveness. *Cyberpsychology, Behavior, and Social Networking*, *14*(11), 643-648.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321-326.

Suler, J. R. (2016). *Psychology of the digital age: Humans become electric*. New York: Cambridge University Press.

Thacker, S., & Griffiths, M. D. (2012). An exploratory study of trolling in online video gaming. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL), 2*(4), 17-33.

The Guardian, (2020). Algorithms on social media need regulation, says UK's AI adviser, (Accessed: 10 December 2020), Available at: theguardian.com/media/2020/feb/04/algorithms-social-media-regulation-uk-ai-adviser-facebook

Twitter for Business (2021), Inappropriate Content, Retrieved from: https://business.twitter.com/en/help/ads-policies/ads-content-policies/inappropriate-content.html

Valkenburg, P. M., & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. *Journal of Adolescent Health*, *48*(2), 121-127.

Williams, R. L., & Cothrel, J. (2000). Four smart ways to run online communities. *MIT Sloan Management Review, 41*(4), 81–91.

Wright, S. (2006). Government-run online discussion fora: Moderation, censorship and the shadow of control. *The British Journal of Politics & International Relations, 8*(4), 550-568.