Examining abuse in online media

Examining abuse in online media

Rahul Sambaraju

Trinity College Dublin

Chris McVittie

Queen Margaret University Edinburgh

Examining abuse in online media

While online spaces and media offer unique possibilities for participating in critical and mundane communication, these also introduce several problems in the form of abuse such as trolling, flaming, or other anti-social behaviour. Social and personality psychologists offer a range of explanations for abusive behaviour online. Here we distinguish between explanations that treat online abuse as readily known and consequently proceed with examining possible causes of such abuse, and those that treat abuse as a situated act of communication worth examining in its own right. This latter, examines how abuse is accomplished, treated, and negotiated in specific online settings. A central advantage of doing so is that the specifics and details of instances of abuse become amenable to examination and consequently open to identification of *in situ* means by which abuse maybe challenged. In taking this approach, online abuse is examined for how it is produced and situated in specific social and interactional settings.

Key words: *Trolling; Online abuse; discourse analysis; social media; online interaction; conversation analysis.*

The rise of many-to-many communication networks initially appeared to offer great promise for a "new public sphere" of citizen journalists and an overthrow of corporate agenda setting in political news gathering, reporting, and framing. In this new public sphere, the people—rather than news media pundits or news corporations—would (re)shape democratic discourse through direct participation (Perlmutter, 2008). Alongside this, participation in online social networks was seen to offer new and practically unlimited opportunities for sharing information and interacting in ways that would transform the essence of everyday social life (Baruah, 2012). However, much research in recent years has suggested that participation has not been the ubiquitous public good that was anticipated: participation can take many forms, positive or negative. In particular, an ever-increasing number of studies have pointed to the incidence of insulting and abusive language found across all forms of online communication (e.g. Binns, 2012; Upadhyay, 2010). A particular case in point is that

of Twitter, one of the most popular social media sites globally, which attracts in excess of three hundred and fifty million active users and has around one billion tweets being posted every five days (Hardaker & McGlashan, 2016). It is especially popular with public figures, who use the site to interact with followers, promote their latest project and gain instant feedback on their activities (Cole, 2015). Yet, this accessibility is a double-edged sword in that any quality or activity of a user can form a target for insults (Bishop, 2013) resulting in individual users potentially receiving considerable abuse (Cole, 2015; Jane, 2014). Indeed, abuse on the site recently reached a level that prompted the company Chief Executive Officer to acknowledge that the company 'sucks at dealing with abuse and trolls on the platform' (Hern, 2015). As with other forms of online media, then, broad participation has been accompanied by a proliferation of abuse made possible through that participation.

Understanding online abuse

The prevalence of abuse found in online settings has unsurprisingly become a topic of interest to many psychologists. For psychologists, however, the challenge is how most usefully to describe such behaviour. Consider the example below, taken from a study of the language often directed against women online (Jane, 2014a).

> Your a ugly, whorish, Slut. I hope someone slaps the fuck out of you and spits in the face … Your nothing more than an easy little cum dumpster.
>
> (Formspring [2011], cited by Jane, 2014a, p.531, original spelling)

In this example, we can clearly see the contributor's use of terms that relate to the addressee's appearance and suggested sexual proclivity. This is combined with an expressed desire that someone violently assaults her. The contribution, then, is designed to offend and it is difficult to read it as other than abusive. Nonetheless, such instances and other similar uses of language found online can be approached in two very different ways, one of which focuses on the message itself and the other of which is more concerned with the interactional consequences of the message. These respective approaches take different issues to be of

primary concern and offer somewhat different understandings of what is to be treated as online abuse.

The first approach focuses on the message as abuse in itself. From this perspective, the terms of the message are self-evidently abusive. Certainly, there is scope for debate over precisely what form of abuse it takes. Much of the abuse found online has been and is described as 'trolling', although writers argue against the use of this term to describe all instances of potentially offensive language. For example, Hardaker (2010, p.224) notes that '"trolling" has become a catch-all term for any number of negatively marked online behaviours', while Dynel (2016, p.317) argues that there is a 'need for clearer definition of trolling vis-à-vis other online communicative behaviors'. Referring to the example above Jane (2014a) rejects the terms 'flaming', 'trolling', and 'cyberbullying', all used interchangeably in previous literature, in favour of the term 'e-bile' to describe language that is 'heavily laced with expletives, profanity and explicit imagery of sexual violence (Jane, 2014b, p.558). Coles and West (2016, p.233) in reviewing the literature suggest that ''trolling' may have multiple, inconsistent and incompatible meanings, depending upon the context in which the term is used and the aims of the person using the term (p.233). There is then no clear consensus as to definitions of terms such as 'trolling' and others[1]. Notwithstanding such debates, however, what we can note for present purposes is that studies conducted within this approach share a common perspective that online abuse is constituted by the message itself.

The second approach also has regard for the message and the descriptions that it includes. Terms that comment negatively upon the intended addressee or that produce apparent threats are of course open to being treated as abusive. Such messages are however also available to be treated and responded to in ways other than as the receipt of direct abuse. For, potentially abusive utterances, like other utterances, can be taken up by a recipient in ways that differ from those anticipated by the sender. For example, studies of ritualised insults or 'insult games' show that those who participate in these activities have available to them a range of possible responses, ranging from the retort of a further insult to the use of 'mis-identifiers' ('mommy') to displays of ostensible gratitude ('thank you') (Sacks, 1992). What this shows is that messages, even potentially extreme ones, do not constitute abuse in

[1] In view of this lack of consensus, we do not attempt here any definition of 'trolling' or other terms. We use the term 'abuse' to delineate a range of behaviours that might be considered to be offensive towards the recipient or recipients. In referring to other studies, we have retained the terms used by the authors.

themselves: insults and abuse are accomplished not in a single utterance but rather over two or more turns in talk. To understand therefore how abuse works whether online or elsewhere, attention turns to the interaction in which the potentially abusive message is produced and examines the consequential relevance of that message for the ensuing exchange. Abuse, within this approach, cannot be treated as self-evident in terms of the originating message and instead becomes an interactional concern for those involved.

These two approaches therefore markedly different ways of examining the abuse found in online settings. Below we consider in detail what they offer to understanding this issue.

Approach 1: abuse as message

From the perspective that online messages framed in negative terms are self-evidently abusive, explanations for that abuse turn from the message itself to the factors that caused the sender to produce the abuse. Attention turns therefore to factors intrinsic to the individual, or factors inherent in how the individual behaves when in contact with others, and how these factors are implicated in the abuse that results in particular settings. We consider these explanations in turn.

Personality traits

One proposed explanation relies on the argument that these problematic actions are motivated by those who are disposed to specific other types of behaviours such as being extraverted or enjoying disagreeable behaviours (Philips & Butt, 2006). Specifically, researchers have identified a set of 'noxious' personality traits, called the 'dark triad' (Buckels, Trapnell, & Paulhus, 2014; Goodboy & Martin, 2015), comprising 1) narcissism, which is a tendency to feel entitled and better than others; 2) Machiavellianism, which refers to a disposition to manipulate others; and 3) psychopathy, which implies lack of remorse and empathy alongside a tendency to engage in disinhibited and egoistical behaviour. These traits have a common feature of including callousness and manipulation as behavioural dispositions (Jones & Figuerdo, 2013). However, more recent arguments call for including a fourth trait, that of sadism, in what becomes a 'dark tetrad' (Van Geel, Goemans, Toprak, & Vedder, 2017).

Van Geel et al (2017) show that enjoying the suffering inflicted on others – sadism – is a significant predictor of cyberbullying in addition to psychopathy.

Some evidence appears consistent with such an explanation. For example, Buckels et al (2014) similarly found that sadism showed substantial associations with trolling behaviours. There is however little evidence to suggest that the prevalent and varied forms of abuse online can simply be traced to the personality traits of those who initiate such abuse, or that a focus on the initiators alone can offer a useful understanding of the interactive nature of such behaviour.

Deindividuation and anonymity

By contrast, other psychologists have sought to explain online abuse in terms of how individuals commonly behave in group settings. A commonly used explanation here is that of 'disinhibition' derived from the concept of 'deindividuation', which proposes that individuals, when acting as part of groups, submerge or immerse their individual identities (Festinger, Pepitone, & Newcombe, 1952). From this perspective, it is argued that such loss of identity allows for minimal restraint and frees individuals to engage in socially unfavoured activities. Zimbardo (1969) offered deindividuation as an explanation for a range of societally problematic or inappropriate actions such as murder, violence, and aggression in 'mobs' or crowds. Postmes et al (2001) argue that deindividuation leads to inhibition of normative influence in online environments. Widyanto and Griffiths (2011) argue that the anonymity afforded by the internet 'removes the threat of confrontation, rejection, and other consequences of behaviour' (p. 15). Users might then disregard norms that are routinely found in face-to-face communication (Arendholz, 2013). Bishop (2013) argues that anonymity allows for 'disinhibition', which then results in higher likelihood for deindividuation, and consequently, the increased possibility of unfavoured activities. Explanations of online behaviour, however, use a thin version of anonymity that refers to issues with identifying name, location, employment, and so on and prevent the identification of users as persons (Lowry et al., 2013). This notion of 'pseudo-anonymity' (Bishop, 2013), where users set-up and use a 'username' of their choice, has been used to explain online behaviour including abuse (Suler, 2004).

For Suler (2004) 'online disinhibition effect' implies that online environments offer possibilities for revealing various forms of 'the self' than offline spaces. He describes six factors that work together with various personality factors in generating minimal inhibition in online spaces: dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and minimization of authority. However, these arguments offer little by way of explanation as to how or when disinhibition might operate online and are built on problematic (see below) theoretical foundations and accompanying evidence.

Group norms and social identities

Another perspective that focuses on individual behaviour differs from those considered above in offering broader group-level explanations, instead of individual-based ones, for online (and offline) behaviours. Derived from social identity theory and the Social Identity Model of Deindividuation Effects (SIDE) (Klein, Spears, & Reicher, 2007; Reicher et al., 1995) this explanation identifies the importance of how different sets of norms become relevant in different group and intergroup conditions. The norms in operation, that is, those that are adhered to or violated, are bound-up with the identities that group membership affords. Tajfel and Turner's social identity theory (1979), treats group-derived identities as salient in settings where there is a ready 'outgroup'. Researchers adopting this perspective argue that online abuse should be viewed not as an irrational outcome resulting from absence of self-regulation but instead as the rational outcome of group-related processes.

Synott, Coulias, and Ioannou (2017) use this approach to examine a very specific instance of online abuse: trolling against the McCann family. The McCann family were at the centre of a controversy related to the unsolved disappearance of 3-year-old Madeleine McCann while her parents were dining at a beach resort in Portugal. The parents' possible culpability in this act was widely held as a reason for popular outrage against them, specifically in the form of trolling online. Trolls here developed group identities, maintained group cohesion, and differentiated themselves from 'pro-McCann' users. These identities brought-up norms related to forms of messages (abusive or not) that can be legitimately posted. For those trolling the McCann family their actions were 'masked' under the guise of seeking justice than as frivolous or mere amusement.

A group-based explanation then argues that online abuse can centrally involve phenomena that are routinely seen in terms of ingroup/outgroup relations. And, it is suggested that online media provide readily available settings for such group processes to operate. For example, Tepper's (1997) examination of trolling activities in a Usenet[2] group, shows that despite Usenet (or other online spaces) being an unmoderated virtual space where users 'come and go' as they please, particular types of groups or communities are routinely found, such as accepted users vs. 'newbies'. Bishop (2014) shows that, trolling directed at users, called 'lurkers', through messages that provoke their participation and then abuse them, prevents them from complete participation in specific online communities. Trolling or online abuse as an activity then can function to maintain communities of users.

Groups found in online media, however, can also mirror those found elsewhere. Flores-Saviaga, Keegan, & Savage (2018) show how 'political trolls' are focused on inviting users identified as 'ideological opponents' into problematic arguments. The insidious form that trolling takes, goes beyond 'lulz' (term for laughter or amusement at another's expense in online settings) and involves suppressing the voices of their ideological or political opponents. Unsurprisingly, abuse directed at members of other groups goes beyond the realm of politics, with online discussions displaying many of the features of prejudice found elsewhere, such as homophobia, misogyny, and racism (Hardaker & McGlashan, 2016). Thus, writers such as Philips (2013) argue that trolling activities need to be examined as part of an extension of dominant cultural activities, such as those of masculinity or heterosexism (see also Cole, 2016). From this perspective, trolling or other forms of online abuse is not unique either in its abusive aspects or that it is online. Rather, it is situated in mundane activities of users and is to be examined as such.

Politeness/impoliteness

One aspect of group-based explanations that the above studies leave relatively unexplored is the issue of how group processes are reflected in specific communicative practices online. In online spaces, social media in particular, the confluence of sociality and the unique interactional environment involves routine norms around communication, such

[2] Usenet is a worldwide distribution discussion system that is free for any user to subscribe.

as those based on 'politeness', alongside those more relevant to technological practices, that is, technology-related norms (Sørensen, 2006, p. 52). The former might attend to generic sociality of communication whereas the latter might range from the use of specific content (emojis and others) to interacting with other users. In her examination of mourning in online spaces, Wagner (2018) argues that norms in social media are constantly in flux and that whatever norms are in play are an outcome of negotiations involving both technological and communicative aspects.

Furthermore, many online environments make readily available tools for evaluating others' behaviours (Watts, 2003), with design features embedding evaluations and assessment in social media activities that offer opportunities to users to indicate im/politeness of specific posts or messages (Zappavigna, 2012). Also, following Culpeper's (2011) work on (im)politeness behaviour in primarily offline social settings, Hopkinson (2014) shows that activities such as flaming and trolling show many of these features. For instance, the role of adopting a persona that does not match with their otherwise 'real' persona not only instantiates deception but also offers strategic advantage. Arendholz (2013) argues that in contrast to offline behaviours, in online spaces the possibility for taking up 'avatars' or 'pseudonyms' affords ready possibilities for impoliteness.

On this view, researchers can examine online abusive behaviour for how it conforms to routine or in-flux norms (Wagner, 2018) of politeness in the 'community of practice' to which the users belong. This line of research allows for an identification of whether some post, message, or online behaviour is impolite, based on certain criteria. However, not all acts of impoliteness constitute abuse or as noted above (Bishop, 2012) what is treated as online abuse need not be impolite or problematic.

These explanations then go beyond a focus on factors internal to the individual and consider online abuse as a more mundane or routine form of activities and socio-cultural arrangements. Nonetheless, what such explanations share with those considered earlier is a focus on individual behaviours and how these are to be judged or evaluated against prevailing social understandings. It is broadly accepted that what constitutes online abusive behaviour, is readily recognisable and that the task for researchers is to explain why individuals engage in such behaviour. And, in the pursuit of this focus, what is omitted is any attention to online abuse in itself, and how it is to be understood as communication.

Approach 2: 'abuse' and interactional outcome

In contrast to the approach outlined above, the second approach to understanding online abuse is concerned not with what might lie behind the message but rather with the act of communication: the message and, if and how it is taken up by the addressee. Focusing on the interactional elements of potential abuse, discursive researchers argue for an examination grounded, first in the form and structure of online communications, and second the orientations and actions of those who produce and respond to such communications (McKinlay & McVittie, 2008). Discourse analysts examine discourse as a topic of study in its own right. The aim is to examine specific elements that go into constructing specific activities and how these activities themselves are constructed as specific types of activities. A core focus of examination is not just constructions but their use in specific occasions to accomplish some social action (Potter, 1996). On this approach, then, using specific lexical items language-users develop specific constructions to accomplish actions such as giving a compliment, raising a query, or insulting. However, these constructions can be taken-up and oriented to in any number of ways. For instance, what might seem like a query can be oriented to as an insult.

In online settings then, users produce posts, messages, or tweets (deploying lexical items and items such as emojis or GIFs as afforded by the online environment) that are designed to accomplish actions such as announcements, compliments, or insults. However, other users who take themselves to be recipients can and do orient to and act on these activities in a range of ways. This means that users can flexibly treat various forms of actions as constituting 'online abuse'. This problem is acute in cases where abuse is racial. Hughey & Daniels (2013) argue that the many-to-many format of online spaces involves a prevalence of racist comments and discourse even on directly unrelated topics. At the same time, the design of several of these spaces implies that explicit racial (or sexist) language cannot be used (Harlow, 2015). A challenge then for the analyst is about how best to examine posts that are while not explicitly racist, carry such implications. Broadly, researchers proceed in one of two ways: one, examine what forms of activities are treated as constituting 'online abuse' or, two, how such activities are accomplished.

Coles and West (2016) examine comments made online in response to an article about trolling, to identify the ways in which posters themselves offered comments about trolls and

trolling. While they adopt a loose definition of trolling as constituting some form of 'disruptive' action, they identify four routine ways of commenting about trolls, which they call 'repertories': 1) that trolls are easily identifiable, 2) nostalgia, 3) vigilantism and 4) that trolls are nasty. They also identify instances where users treat some trolling as 'acceptable', such as where a user is seen to be 'trolling the trolls'. Petyko (2018) examines instances of political trolling in left-wing Hungarian blogs. The examination here focuses on how ascribing specific motives to fellow posters allowed for claims that those were trolls. Such findings demonstrate that, what counts as trolling and either as problematic or not is bound-up with the context within which these activities take place.

Whatever may constitute online abuse does take place in specific instances on a specific platform with specific users (inclusive of bots[3]). In that these are situated activities. Conversation analysts and researchers who examine interaction offer a profoundly alternative view of such activities (Sacks, 1995; Housley et al, 2017). These researchers focus on the role of interactional features of online activities and the consequences these have on how activities are shaped, taken-up, and responded to. Numerous studies from the 1990s onwards (McKinlay, Procter, Masting, Woodburn, & Arnott, 1994) have shown that online talk is amenable to analysis that applies key elements of conversation analysis. Paulus and colleagues (Paulus, Warren & Lester, 2016) report that conversation analysis has become an increasingly popular method of analyzing turn-design and sequence organization in online talk, allowing for detailed examination of how participants accomplish social actions in these contexts. Housley et al (2017) argue that ethnomethodological approaches inclusive of conversation analytic and membership categorization analytic techniques can be of direct relevance to study social media interactions. They argue that tweets on Twitter can be examined as units of analysis for the actions that these accomplish.

While there are concerns over the feasibility of using techniques that focus on sequential organization for analysis of online communication (Greiffenhagen & Watson, 2005 as cited in Meredith, 2017): these posts will not incorporate many of the elements found in speech, such as intonation, hesitations, or speed of delivery. The resulting conversations are thus likely to diverge considerably from naturally occurring face-to-face interactions. Conversation analytic researchers however focus on how posts or tweets on the internet are

[3] A programmed profile or user-account that posts pre-programmed messages on online spaces.

designed to 'do' some social action in their content and sequential organization (Meredith, 2017). Online interaction in the form of talk or other symbols is examined as a social practice. There is then a growing focus to treat online interactions as 'social practices in their own right' (Lamerichs & te Molder, 2003). For instance, researchers recognize that sequential actions might be disrupted in online interactions, which however is not treated as problematic by users themselves (Berglund, 2009). Users then can and do work with and around the technological affordances that various online platforms provide (Gibson, 1979; Hutchby, 2001).

In line with such approaches, Housley et al., (2017b) examine how a controversial celebrity Katie Hopkins employed specific membership categories and devices, to develop antagonistic inferences. More recently, Procter et al (2019) show that some forms of responses that counter potentially racist, sexist, or homophobic tweets can work to suppress the spread of 'cyber-hate'. However, the tweets and the responses are limited to specific forms of problematic issues that for the authors constitute 'cyber-hate'. These studies however do not directly examine how instances of 'online abuse' themselves are accomplished *in situ*.

Jenks (2019) addresses a highly pertinent issue: any comment or post from a user can potentially be treated as an attempt at trolling or a form of 'online abuse'. It then is of interest to examine how and in what ways the recipients or fellow users might treat any specific instance. Using conversation analysis, Jenks offers a conversation analytic examination of how users orient to, treat, and negotiate what is to count as trolling on a discussion board for migrants living in Hong Kong. The analysis shows the central role of how users orient to and use as a resource 'floor spaces' in identifying and negotiating trolls and trolling behaviours. By floor spaces, Jenks refers to the discursive space that allows those participating to identify and develop claims of appropriate or inappropriate conduct. In other words, floor space is a version of the context within which an ongoing interaction is taking place. Jenks (2019) identifies various ways in which users could make relevant the properties and features of the discussion forum in identifying and negotiation whether a comment or post is an instance of trolling.

The above studies then show that instances of 'online abuse' can be closely examined for how these are produced and in turn taken-up by other users. Hardaker (2015) proposes that researchers ought to examine how those who are the recipients (intended or otherwise)

of 'online abuse' respond to such comments or posts. Moreover, as Proctor et al (2019) report, these can be used to minimize, if not mitigate, specific forms of widely recognized hate-talk. A close examination of those specific instances (see below) shows the active, open, and collaborative aspects of activities identified as 'online abuse'.

Approaches 1 and 2: a comparison

In order to illustrate the differences between the two approaches to understanding abuse that we have outlined above, let us consider a particular example of potentially abusive talk. This talk comprises an exchange taken from the Twitter feed of one individual, James Blunt, a well-known UK singer-songwriter who is recognised for receiving a considerable volume of potential abuse on twitter and for the responses that he produces. To ease readability of the exchange, we present the turns in chronological order below.

1	Tweet	@ooliviae: @James Blunt is the rudest cunt on this earth, I fucking hate him.
2	Retweet and response	James Blunt @JamesBlunt 17 Sep 2015 U're just a jealous runner-up in the Rudest Cunt Competition.

(from https://twitter.com/, cited in McVittie & Sambaraju, 2019)

Twitter as a social media platform offers unique affordances, such as that tweets posted can be replied to directly or replied along with re-posting them, called retweeting (see Calvin et al., 2015). Retweeting then shows to an overseeing audience the reply along with the original tweet. This can attend to performative aspects of using Twitter or other online media.

The exchange seen above begins with a description of Blunt that he retweets in his response. This tweet depicts Blunt and his behaviour in extreme terms, describing him as being 'the rudest cunt on this earth'. It is immediately followed by a statement of ooliviae's personal feelings towards Blunt, that he/she 'fucking hate(s) him'. This turn then is set out in terms that are potentially highly derogatory and insulting.

From the perspective of the first approach outlined in this paper, these features of the

tweet would be sufficient for it to constitute abuse in this online context. Even allowing for online norms of behaviour, the behaviour can readily be viewed as impolite. There would no need to consider the tweet itself further, except perhaps to determine whether it comes within a specific category of online abuse, however defined. Interest then would turn to the question of why ooliviae posted this tweet, in particular whether it was motivated by his/her personality traits, the anonymity available in this context, the group/identity processes operating, or some combination of these. Blunt's response thus comes to be seen as a means of dealing with one instance of the abuse that he commonly receives in this setting.

By contrast, within the second approach considered here the tweet is viewed as potentially abusive instead of necessarily abusive: whether or not it constitutes abuse depends on how Blunt orients to it as seen in his response. And, turning attention to the exchange as a whole, it becomes apparent that Blunt does not treat this as abuse in his response. Instead, he takes up two elements of the tweet in responding to ooliviae. First, he ascribes to oolivae membership of the same membership category, 'rudest cunt', that he/she introduced, and constructs him/her as a less successful member than he himself is in describing oolivae as a 'runner-up' in a membership-based 'competition'. Second, Blunt uses this ascription to impute to oolivae a motivation for posting the tweet, suggesting that oolivae is 'jealous' on the grounds on Blunt's greater achievement. In adopting this framing, Blunt's response orients to the tweet as indicating nothing more than a display of oolivae's disgruntlement following a lack of personal success. In interactional terms, abuse is not accomplished over the two turns and the potentially derogatory tweet becomes no more than a failed attempt to abuse Blunt. Online abuse, much like any other action is then better examined as an orientation and concern for participants themselves.

Conclusions

In the present paper, we focused on contemporary explanations for online abuse from a range of broadly psychological perspectives. A key difference lies in whether approaches to such behaviour proceed with an *a priori* classification, categorization, or knowing of abuse, or whether abuse is treated as an accomplishment *in situ*. We argue that the former treat abuse as readily identifiable and therefore an examination of the causes or factors involved in producing abuse or to treat abuse as a concern for users themselves. The latter treats abuse

as an accomplishment that involves various participants who can variously construct and orient to messages and posts as abusive or not.

.

We specifically argue for the latter approach to examine trolling or other forms of online abuse that treats it as an interactional accomplishment. Specific implications follow from this. First, this mitigates concerns over how best to code some text or online practice, including GIFs and memes as abuse. Second, it allows for examining how potential abuse is justified, normalized, or challenged. Third, it allows for observing, describing, and examining what it means for users to engage in these activities in various roles. Together then this approach eschews an *a priori* classification of messages as trolling, flaming, or abuse.

.

Future research can then usefully proceed along two lines: first, research can examine discourse about trolling, flaming, or abuse. This would be an examination of how various posts are constructed as constituting or not online abuse. Second, research can examine instances of what could potentially be taken as trolling, flaming, or abuse. This would be an examination of how these posts are designed to be read as abusive but are then variously oriented to and negotiated. A focus on these activities of users will engage with how specific identity constructions (McKinlay & McVittie, 2011), category memberships (Giles, 2017) or other social psychological concerns are used by users to accomplish these activities.

In these ways, approaches that prioritize the discourse about or involving online abuse offer researchers tools to see these activities from the perspective of those involved. These tools have the potential to offer a radically different view of online activities, including those of trolling, flaming, and verbal abuse: these are accomplishments whose meaning and relevance are open to negotiation.

References

Arendholz, J. (2013). *Appropriate Online Behavior: A Pragmatic Analysis of Message Board Relations*. Amsterdam: John Benjamins.

Baruah, T. D. (2012). Effectiveness of Social Media as a tool of communication and its potential for technology enabled connections: A micro-level study. *International Journal of Scientific and Research Publications*, *2*(5), 1-10.

Berglund, T.Ö. (2009) Disrupted turn adjacency and coherence maintenance in instant messaging conversations. *Language@Internet 6(2).* Available at www.languageatinternet.de/ articles/2009/2106/Berglund.pdf/.

Binns, A. (2012). DON'T FEED THE TROLLS! Managing troublemakers in magazines' online communities. *Journalism Practice, 6*, 547–562. doi:10.1080/17512786.2011.648988.

Bishop, J. (2013). The effect of deindividuation of the internet troller on criminal procedure implementation: An interview with a hater. *International Journal of Cyber Criminology,* 7, 28-48

Bishop, J. (2014) 'The psychology of trolling and lurking: the role of defriending and gamification for increasing participation in online communities using seductive narratives', in Li, H. (Ed.) *Virtual Community Participation and Motivation: Cross-Disciplinary Theories.* IGI Global, Hershey, PA.

Buckels, E., Trapnell, P. & Paulhus, D. (2014). Trolls just want to have fun. *Personality and Individual Differences, 67*, 97-102. Doi: 10.1016/j.paid.2014.01.016

Calvin, A.J., Bellmore, A., Xu, J. & Zhu, X. (2015) #bully: Uses of hashtags in posts about bullying on Twitter. *Journal of School Violence, 14:1*, 133-153, DOI: 24 10.1080/15388220.2014.966828

Coles, B. & West, A. (2016). Trolling the trolls: online forum users constructions of the nature and properties of trolling. *Computers in Human Behaviour*, 60, 233-244

Cole, K. (2015). "It's like she's eager to be verbally abused": Twitter, trolls, and (en) gendering disciplinary rhetoric. *Feminist Media Studies, 15*, 356–358. doi:10. 1080/14680777.2015.1008750.

Costa, P., & McCrae, R. (1992). Revised NEO Personality Inventory (NEOPI–R) and Five Factor Inventory (NEO–FFI) professional manual. Odessa, FL: *Psychological Assessment Resources.*

Festinger, L., Pepitone, A. & Newcomb, T. (1952) Some Consequences of De-Individuation in a Group. *Journal of Abnormal and Social Psychology*, 47, 382- 389. DOI: https://doi.org/10.1037/h0057906

Flores-Saviaga C, Keegan B, & Savage S (2018) Mobilizing the trump train: Understanding collective action in a political trolling community. In: ICWSM'18, URL https://www.aaai.org/ocs/index.php/ ICWSM/ICWSM18/paper/view/17877/16999

Gibson, J. (1979). *The ecological approach to perception*. London: Houghton Mifflin.

Giles, D. (2017). How do fan and celebrity identities become established on Twitter? A study of 'social media natives' and their followers, *Celebrity Studies*, 8, 445-460, DOI: 10.1080/19392397.2017.1305911

Goodboy, A. & Martin, M. (2015). The personality profile of a cyberbully: Examining the dark triad. *Computers in Human Behavior*, 49, 1–4.1

Greiffenhagen, C. & Watson, R. (2009). Visual repairables: analysing the work of repair in human–computer interaction. *Visual Communication,* 8, 65–90. https://doi.org/10.1177/1470357208099148

Hardaker, C. (2015). "I refuse to respond to this obvious troll": an overview of responses to (perceived) trolling. *Corpora*, 10, 201-229. Doi: https://doi.org/10.3366/cor.2015.0074

Hardaker, C., & McGlashan, M. (2016). ''Real men don't hate women'': Twitter rape threats and group identity. *Journal of Pragmatics,* 91, 80–93. doi:10.1016/j.pragma.2015.11.005.

Hern, A. (2019). Twitter CEO: We suck at dealing with trolls and abuse. Retrieved 25 September 2019, from https://www.theguardian.com/technology/2015/feb/05/twitter-ceo-we-suck-dealing-with-trolls-abuse

Hopkinson, C. (2014). 'Face Effects of Verbal Antagonism in Online Discussions,' *Brno Studies in English* 40, 65 - 87.

Housley W. et al. (2017a) Digitizing Sacks? Approaching social media as data. *Qualitative Research 17,* 627–644.

Housley, W. et al. (2017b). Membership Categorisation and Antagonistic Twitter Formulations. *Discourse & Communication, 11,* 567-590.

Hughey, M. & Daniels, J. (2013). Racist comments at online news sites: A methodological dilemma for discourse analysis. *Media, Culture & Society,* 35, 332–347. doi:10.1177/0163443712472089.

Jane, E. A. (2014a). "Your a ugly, whorish, slut": understanding E-bile. *Feminist Media Studies*, *14*(4), 531-546.

Jane, E. (2014b), 'Back to the Kitchen, Cunt': Speaking the Unspeakable About Online Misogyny, *Continuum: Journal of Media and Cultural Studies*, 28, 558–70.

Jenks, C. (2019). Talking trolls into existence: On the floor management of trolling in online forums. *Journal of Pragmatics, 143*, 54-64. Doi: https://doi.org/10.1016/j.pragma.2019.02.006

Jones, D., & Figueredo, A. (2013). The core of darkness: Uncovering the heart of the Dark Triad. *European Journal of Personality, 27*, 521–531.

Klein, O., Spears, R., & Reicher, S. (2007). Social Identity Performance: Extending the Strategic Side of SIDE. *Personality and Social Psychology Review*, 11(1), 28–45. https://doi.org/10.1177/1088868306294588

Lamerichs, J. & te Molder. H. (2003). Computer-mediated communication: from a cognitive to a discursive model. *New Media & Society, 5*, 451-473. Doi: 10.1177/146144480354001

Lowry, P., 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,* 962–986

McCosker, A. (2014). Trolling as provocation: YouTube's agonistic publics. Convergence: *The International Journal of Research into New Media Technologies, 20,* 201-217. doi:1354856513501413.

McKinlay, A., & McVittie, C. (2008). *Social Psychology & Discourse*. Sussex: Wiley-Blackwell.

McKinlay, A., & McVittie, C. (2011).*Identities in Context: Individuals and discourse in action.* Sussex: Wiley-Blackwell.

McKinlay, A. Procter, R. Masting, O. Woodburn, R. Arnott, J. (1995). Studies of turn-taking in computer-mediated communications. *Interactional Computing*, 6, 151-171,DOI; 10.1016/0953-5438(94)90022-1

McVittie, C., & Sambaraju, R. (2019, June). 'I love James Blunt as much as I love herpes' – 'I love that you're not ashamed to admit you have both': emotion talk in insult-retort sequences on Twitter. In C. McVittie (Chair) *Examining the language of Cognitions, Sensations, and Emotions: The discursive construction of mental states in different media.* Paper presented at the 6th Annual Meeting of the American Psychological Association Division 5 Society for Qualitative Inquiry in Psychology, Boston, MA.

Meredith, J. (2017). Analysing technological affordances of online interactions using conversation analysis. *Journal of Pragmatics, 115*, 42–55. https://doi.org/10.1016/j.pragma.2017.03.001

Paulus T, Warren A and Lester JN (2016) Applying conversation analysis methods to online talk: A literature review. *Discourse, Context & Media*, 12: 1–10.

Perlmutter, D. (2008). *Blogwars.* Oxford: Oxford University Press.

Petyko, M. (2018). The motives attributed to trolls in metapragmatic comments on three Hungarian left-wing political blogs. *Pragmatics, 28*, 391-416. https://doi.org/10.1075/prag.17007.pet

Phillips, J., & Butt, S. (2006). Personality and self-reported use of mobile phones for games. *CyberPsychology & Behavior*, 9, 753–758. http://dx.doi.org/10.1089/ cpb.2006.9.753.

Phillips. W. (2015). *This Is Why We Can't Have Nice Things: Mapping the Relationship Between Online Trolling and Mainstream Culture*. Harvard: MIT Press

Postmes, T., Spears, R., Sakhel, K. & DeGroot, D. (2001). Social influence in computer-mediated communication: The effect of anonymity on group behavior. *Personality and Social Psychology Bulletin, 27*, 1243-1254.

Potter, J. (1996). *Representing Reality: Discourse, Rhetoric, and Social Construction*. London: Sage

Proctor, R., et al (2019). A Study of Cyber Hate on Twitter with Implications for Social Media Governance Strategies. In *Proceedings Conference on Truth and Trust Online*. Accessed on September 15, 2019 from: https://arxiv.org/abs/1908.11732

Reicher, S., Spears, R. & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, 6, 161-198. Doi: 10.1080/14792779443000049

Sacks, H. (1995). *Lectures on Conversation*. Blackwell: Oxford

Sørensen, K. (2006). Domestication: The enactment of technology. In Berker, T., Hartmann, M., Punie, Y., Ward, K. J. (Eds.), *Domestication of media and technology*, pp: 40–60. New York, NY: Open University Press.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7, 321–326. http://dx.doi.org/10.1089/1094931041291295.

Synnott, J. Coulisas., A. & Ionnaa, M. (2017). Online trolling: The case of Madeleine McCann. *Computers in Human Behavior, 71,* 70-78

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33-47). Monterey, CA: Brooks/Cole

Tepper, M. (1997). Usenet communities and the cultural politics of information. In D. Porter (Ed.), *Internet culture* (pp. 39–54). New York: Routledge.

Upadhyay, S. R. (2010). Identity and impoliteness in computer-mediated reader responses. *Journal of Politeness Research,* 6, 105–127.

Van Geel, M., Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences,* 106, 231–235. https://doi.org/10.1016/j. paid.2016.10.063

Wagner, A. (2018). Do not Click "Like" When Somebody has Died: The Role of Norms for Mourning Practices in Social Media. *Social Media + Society,* 1–11

Watts, J. (2003) *Politeness*. Cambridge: Cambridge University Press

Widyanto, L. & Griffiths, M. (2011). An empirical study of problematic internet use and self-Esteem. *International Journal of Cyber Behavior Psychology and Learning*, 1, 13-24. Doi: 10.4018/ijcbpl.2011010102

Zappavigna. M. (2012). *Discourse of Twitter and Social Media: How we use Language to Create Affiliation on the Web.* Continuum, London: UK.

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation,* (pp. 237–309). Lincoln: University of Nebraska Press.